FROM PHARMACOVIGILANCE TO CLINICAL CARE OPTIMIZATION

Leo Anthony Celi,[1] Edward Moseley,[2] Christopher Moses,[3] Padhraig Ryan,[4] Melek Somai,[5] David Stone,[6] and Kai-ou Tang[7]

Abstract

In order to ensure the continued, safe administration of pharmaceuticals, particularly those agents that have been recently introduced into the market, there is a need for improved surveillance after product release. This is particularly so because drugs are used by a variety of patients whose particular characteristics may not have been fully captured in the original market approval studies. Even well-conducted, randomized controlled trials are likely to have excluded a large proportion of individuals because of any number of issues. The digitization of medical care, which yields rich and accessible drug data amenable to analytic techniques, provides an opportunity to capture the required information via observational studies. We propose the development of an open, accessible database containing properly de-identified data, to provide the substrate for the required improvement in pharmacovigilance. A range of stakeholders could use this to identify delayed and low-frequency adverse events. Moreover, its power as a research tool could extend to the detection of complex interactions, potential novel uses, and subtle subpopulation effects. This far-reaching potential is demonstrated by our experience with the open Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) intensive care unit database. The new database could also inform the development of objective, robust clinical practice guidelines. Careful systematization and deliberate standardization of a fully digitized pharmacovigilance process is likely to save both time and resources for healthcare in general.

Limitations of Clinical Trials

In the quest for the safe and effective use of pharmaceuticals, there is a pressing need for improved surveillance to reduce the risk of preventable morbidity and mortality after products are released. Clearly, such pharmacovigilance represents a form of care optimization as postrelease safety cannot be maximally assured even by excellent premarket studies. This is particularly so because these agents are prescribed to patient populations that are often not completely represented in the original studies required for market approval. Randomized controlled trials (RCTs), the gold standard for defining drug efficacy and safety profiles, are often of inadequate duration, are too narrow in indication and scope, and consist of relatively few test subjects.[1] Therefore, RCTs are ill-equipped to fully unravel heterogeneity of treatment

[1]*Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.*
[2]*Division of Vaccine Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts.*
[3]*Smart Scheduling, Inc., Cambridge, Massachusetts.*
[4]*Trinity College Dublin, Dublin, Ireland.*
[5]*Department of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, Massachusetts.*
[6]*Departments of Anesthesiology and Neurosurgery and the Center for Wireless Health, University of Virginia School of Medicine, Charlottesville, Virginia.*
[7]*Johns Hopkins School of Medicine, Baltimore, Maryland.*
All authors contributed equally to this work.

effect.[2] In order to test the efficacy of an intervention, RCT participants are specifically selected to form a homogenous cohort, necessitating exclusion of patients whose comorbidities might modulate the treatment effect. In highly cited clinical trials, around 40% of identified patients with the condition under consideration are not enrolled, principally because of restrictive eligibility criteria.[3] One review revealed that only 35% of these studies reported enough information to categorize the reasons for nonenrollment. For example, the efficacy of warfarin for stroke prevention in patients with atrial fibrillation is established as grade 1A evidence, but many potential participants were excluded from these trials, frequently because of the presence of comorbidities.[4] Consequently, there is insufficient understanding of the effectiveness of many interventions in those groups who were underrepresented in the clinical trials.

Other weaknesses of these trials include their limited ability to identify low-frequency adverse events or long-term side effects, as well as wide variations in the choice of comparators, end points, duration, and size.[5] Furthermore, drug–drug interactions associated with complex comorbidities are rarely studied and ultimately impossible to identify within this study design. Results frequently are reported and interpreted as a point estimate, and variability may be viewed as unhelpful measurement error.[6] However, this variability may indicate markedly diverse treatment responses between different patient subpopulations. Another problem is the lack of accessibility of postmarket studies. As of January 23, 2014, there were 159,814 postmarket studies registered with the National Institutes of Health, but the results of fewer than 10% of these had been made publicly available[7] despite evidence suggesting that many patients are willing to share their healthcare data for research purposes.[8] Furthermore, over one-third of drug approvals between 2005 and 2012 were based on a single trial. For the most part, there is simply insufficient evidence with which to draw significant, clinically relevant conclusions. These issues warrant the broadening of postmarket pharmacovigilance to formulate comprehensive drug safety and efficacy profiles.

"THESE ISSUES WARRANT THE BROADENING OF POSTMARKET PHARMACOVIGILANCE TO FORMULATE COMPREHENSIVE DRUG SAFETY AND EFFICACY PROFILES."

Observational and Open Data Approaches to Pharmacovigilance

Clinicians have become cognizant of the aforementioned limitations of RCTs and the need to tailor treatment protocols to the demographic characteristics and comorbidities of patients. The requirement for a system of continuous learning to address these knowledge gaps is increasingly acknowledged.[9] Given the limitations of clinical trials, postmarket observational studies play an essential role in assessing the true risk profile of drugs, devices, and interventions. While the use of observational studies as a complement to RCTs remains debated, we contend that the increasing richness and accessibility of routine health records data, in conjunction with advanced analytical techniques, strengthen the potential complementary role of observational evidence.

The recent body of work from our laboratory supports this position. Over the past decade, the Laboratory of Computational Physiology, Beth Israel Deaconess Medical Center (BIDMC), and Philips Healthcare, with support from the National Institute of Biomedical Imaging and Bioinformatics, have partnered to build, maintain, and analyze the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database.[10,11] The precursor of MIMIC was originally developed in the 1980s by Dr. Roger Mark for the analysis of cardiac dysrhythmias. At that time, the norm was to privately create closed databases, but Dr. Mark wisely determined that an open model would accelerate and generally improve learning from the clinical material. This shared data have been extremely successful in stimulating research interest and beneficial competition, as well as serving as a resource for testing algorithms. The initial success led to the development of ongoing open databases, including MIMIC, now in its second version. Notably, this success has depended on and benefited from collaboration among a variety of expert users, including clinicians, researchers, and technical specialists, as well as endorsement from the hospital leadership. This public-access database, which now holds clinical data from over 60,000 stays in BIDMC intensive care units (ICUs), has been meticulously de-identified and is freely shared online with the research community via PhysioNet (www.physionet.org/). The readers are hereby directed to the appendix section and a review article that provides a detailed description of the MIMIC database and how it came about.[12]

Analysis of MIMIC has deepened our understanding of heterogeneity between different subpopulations, for example, in the treatment effect of red blood cell transfusion.[13] By day 3 of treatment in an ICU, approximately 95% of patients have abnormally low hemoglobin levels, and red blood cell transfusions are frequently administered. These clinical practice patterns have traditionally rested on personal intuition and the habits of care givers rather than evidence. The appropriate indications and dosage of this transfusion therapy are uncertain. Studies reporting mortality and morbidity outcomes have variably reported the effects as protective, harmful, or neutral. How can such an important issue be approached and resolved?

The MIMIC investigators hypothesized that patients' ages, comorbidities, and prior clinical interventions could modulate the effect of transfusion on clinical outcomes. Although the aggregate effect of transfusion on the entire cohort appeared to be neutral, there were important distinctions between various subpopulations. Younger transfusion recipients had higher mortality rates than control patients after adjustment for propensity to receive transfusion whereas older recipients had lower 30-day and 1-year mortality rates than controls. Among patients with heart disease, the outcomes were worse for those who underwent transfusion. This study countered the perception that transfusion in critically ill patients may be uniformly harmful or beneficial.

Another study used MIMIC to uncover association between *ex-ante* use of selective serotonin reuptake inhibitors and mortality in the ICU for certain patient subsets.[14] This type of research that looks into the effect of prior use of particular medications on outcomes during the course of ensuing conditions such as critical illness is an example of clinical questions best addressed by the targeted analysis of large databases. A noteworthy strength of this study was its use of directed acyclic graphs to display the proposed causal pathways that were captured in the analysis. The analytical techniques and findings of these studies may serve as examples of the detail of information provided by retrospective data analyses that are not possible with traditional RCTs.

Pharmacovigilance can be thought of as a special use case or subset of comparative effectiveness research (CER): pharmacovigilance identifies outcomes that are interpreted as "adverse" (rather than effective) because of a drug, drug–drug interaction, or drug–disease interaction. For decades, the Food and Drug Administration (FDA) has relied on voluntary reporting of adverse events by healthcare practitioners and patients. This arrangement has demonstrated many limitations, and awareness has been raised of its shortcomings as representing passive or "reactive surveillance."[15] To meet rising concerns, large-scale projects have been established in recent years to develop active surveillance tools through the use of routinely collected electronic health information. Mini-Sentinel, a pilot project of the FDA's Sentinel Initiative, uses a distributed data approach with a centralized portal to collect aggregated de-identified results and to distribute manually coded packages.[16,17]

There have been efforts to leverage the "big data" potential of Mini-Sentinel, but these had notable limitations, in part because of the closed nature of the data. Seeking to obtain insight into the risks associated with dabigatran, one of the new generation of anticoagulant medications, the FDA deployed Mini-Sentinel to compare the incidence of hemorrhage between patients on dabigatran and patients on the more widely used and much cheaper warfarin.[18] The advisory announcement that ensued from the study stated that bleeding was 1.8–3.0 times greater for warfarin than for dabigatran. However, the confidence intervals were not publicized, and more importantly, the analysis was not adjusted for age, gender, or any clinical differences between the patient populations. Age and gender are established risk factors for hemorrhage from anticoagulant use, and the indication for anticoagulation may have differed significantly between the two groups. Months after the advisory announcement, the FDA released pages of bar charts depicting the risk of hemorrhage stratified into age and gender categories. But no summary analysis was presented to adjust for age and gender simultaneously, or for any clinical factors. Over the years after the release of dabigatran, the FDA has failed to leverage the potential of Mini-Sentinel to clarify the associated risk of bleeding.

"A RECENT ANNOUNCEMENT BY THE FDA SUGGESTS THAT HELP IS NEEDED TO FULLY ACHIEVE THE VISION OF AN ACTIVE SURVEILLANCE PLATFORM FOR PHARMACEUTICALS."

A recent announcement by the FDA suggests that help is needed to fully achieve the vision of an active surveillance platform for pharmaceuticals. The FDA is seeking partners to develop and monitor a database of electronic health records (EHRs) with the goal of expanding its Sentinel Initiative in order to detect adverse events more reliably and to identify predictive risk factors.[19]

There are examples of related initiatives that have been successful. Maguire and Dhar had previously employed data mining on a large scale using a medical and pharmacy claims database to determine patterns of cost and quality.[20] Another example comes from the Mayo Clinic, which partnered with Optum Labs in analyzing administrative claims and clinical data from millions of patients.[21] The research partnership has performed numerous studies, including CER that involve cardiac, diabetes, and anticoagulant medications.

While a predictive capability is an ideal goal of the system we describe in this article, we would initially conceive pharmacovigilance as an observational process. We propose the target database to be the EHRs of *all* patients that contain medication data as well as all other documented clinical information (e.g., vital signs, laboratory results, and provider notes) (Fig. 1). While this kind of large, de-identified clinical database does not currently exist, the concept is quickly coming to fruition through such efforts as that of PCORI.[22] Recently, two of the authors of this work reported on the use of such a database for real-time decision support called dynamic clinical data mining (DCDM) via the integration of medical big data, search engines, and EHRs (Fig. 2).[23]

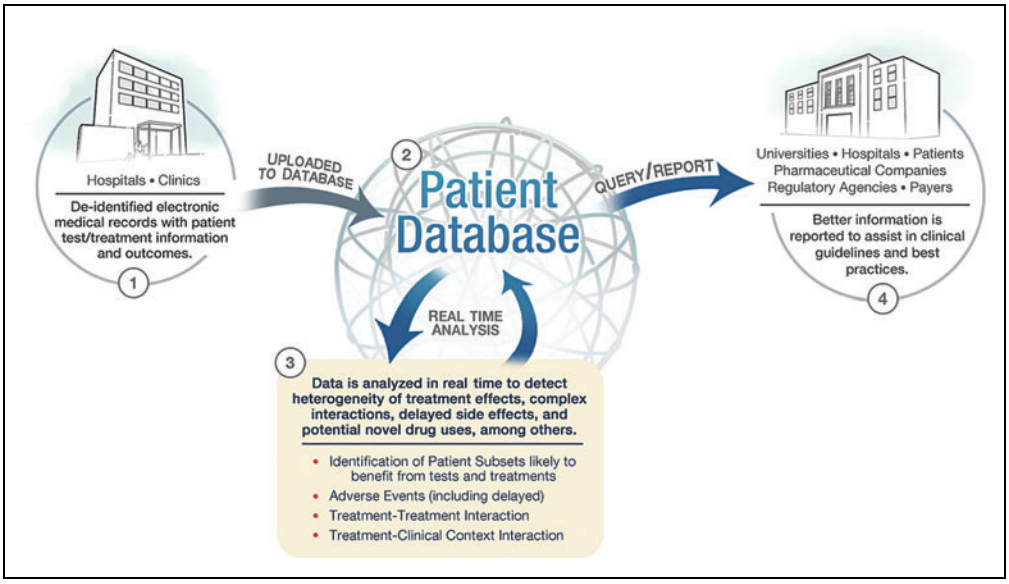

FIG. 1. Clinical care optimization: a big data model for efficient targeting of tests and treatments and vigilance for adverse events.

To illustrate an example, let us consider the possibility of acute kidney injury as an unforeseen delayed adverse effect of a new drug within a patient subset of a certain demographic and with a specific comorbidity. How could such an adverse effect be captured in an EHR-based pharmacovigilance system? Our previously mentioned concept of DCDM performs automatic searches of the universal de-identified EHR database to determine prior outcomes and treatments in similar patients in order to provide individualized decision support for both the provider and the patient. An EHR-based pharmacovigilance system will require an automated data mining tool, analogous to the use of search engines in DCDM, that would identify patterns of developing and completed anomalies in populations. This tool would then report the observation to the appropriate group who would then conduct a more rigorous analysis. The big data design would allow the

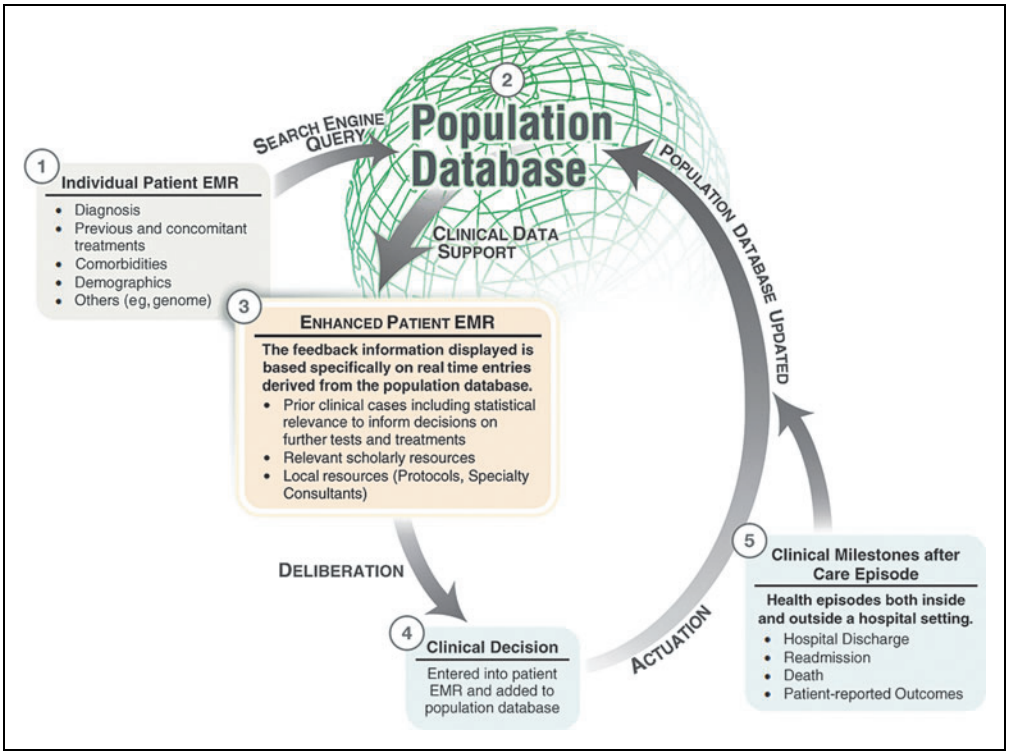

FIG. 2. Dynamic clinical data mining (reproduced with permission from *JMIR Medical Informatics*).

system to detect adverse effects caused by a drug, drug–drug interaction, and drug–disease interaction in populations not well studied in RCTs such as the elderly and particular ethnic and socioeconomic groups. So in our example, the pharmacovigilance system would be sensitive enough to note a rising incidence of acute kidney injury (predefined in the data mining tool) even if it only occurred in Hispanic women over 80 years old with a certain comorbidity or who are concurrently on another specific drug.

Application of MIMIC to Pharmacovigilance

The observational method of research that the FDA intends to implement has already been accomplished in academic as well as industrial institutions. In view of this established success, it seems counterproductive to restrict access to a database for pharmacovigilance to just a few stakeholders. The MIMIC database, with its open source nature, may serve as a model if the FDA seeks to maximize the benefits of its initiative. The primary objective of this new database would not be limited to pharmacovigilance, but will encompass a wide variety of applications within healthcare. This approach would reduce the risks of postmarket health crises by monitoring events in real time. But its use should extend to the more general optimization of clinical care, including CER, the detection of complex interactions, potential novel uses, and subtle subpopulation effects.

Given the labor-intensive nature of retrospective analyses as currently conducted on MIMIC and similar databases, the task of creating comprehensive profiles for each treatment or intervention may be considered prohibitively difficult without automation of many of the processes involved. One way to surmount this barrier is to adopt common data storage methods and advanced machine learning algorithms. A common data storage method would facilitate seamless integration of clinical datasets for analyses that would yield results in almost real time, rather than the several months it takes to perform these analyses by even the most advanced teams of clinicians and data scientists. Like any technology in its infancy, but especially one with the ambition of creating robust clinical guidelines, it is important to keep the database open so that other methodologies may be tested on the same data to further develop, confirm, or contest prior findings.

The success of MIMIC signals that a large-scale, open-access healthcare database is feasible, would be widely used, and would contribute significantly to the advancement of medical knowledge. Researchers of various professional backgrounds are using MIMIC for analyses that include the assessment of outcomes in patients prescribed specific medications or nonpharmacologic interventions, in a wide variety of clinical contexts, and for patients with different permutations of comorbidities. The history of this database shows that its open nature accelerates the creation of new knowledge and allows for the rigorous scrutiny of prior research findings. Over 1,000 investigators from over 32 countries have unrestricted access to the de-identified clinical data based on data use agreements (Roger Mark, personal communication).

In practice, there are barriers to implementing this kind of system. Indeed, the system may experience a "collective action problem," a situation where many stakeholders stand to benefit from an action, but the cost for each individual stakeholder renders it difficult to undertake alone. This may be relevant to the production of a large-scale database if providers are reluctant to contribute their own data. However, a solution to overcome a similar problem was developed by a national leadership meeting at the Department of Health and Human Services in 2011, involving leaders from health plans, purchasers, hospitals, physician specialty groups, and the pharmaceutical industry.[24] Although the focus was on clinical registries, the recommendations can serve as helpful guideposts relevant for the proposed database. For example, payment mechanisms that mandate or incentivize participation in the open database could be devised by the Center for Medicare and Medicaid Services and private health plans. These groups could also facilitate a standard data infrastructure that supports sharing and querying of clinical data. Physicians, medical specialty societies, and research institutions could inform the process of selecting the critical data elements to incorporate. By instilling appropriate financial and intrinsic incentives for key stakeholders, implementing this open database in a collaborative manner would be pragmatic and feasible for all parties involved.

> "BECAUSE OF THE ACTIVE NATURE OF THIS APPROACH, IT IS CLEAR THAT CLINICALLY RELEVANT INFORMATION WOULD BE GENERATED AT A RAPID RATE AS NEW DATA ARE COLLECTED AND ANALYZED."

Heterogeneity of Treatment Effect and Clinical Practice Guidelines

An active pharmacovigilance system could and should extend beyond the basic querying of large data sets. The use of machine learning techniques to mine EHRs in real time can uncover complex relationships between pharmaceuticals, devices, clinical context, comorbidities, and patient demographic factors. The identification of these effects would provide clinicians with the information necessary to more

efficiently categorize patients as regards who will benefit from, be harmed by, or will neither benefit nor be harmed by the use of particular interventions. Because of the active nature of this approach, it is clear that clinically relevant information would be generated at a rapid rate as new data are collected and analyzed. When this process deploys techniques such as instrumental variables, and appropriate use of training and testing datasets, it can have a high degree of validity. In some cases, it would be appropriate to validate hypotheses in a prospective, experimental but pragmatic manner. The associated cost would be modest as compared to traditional RCTs given the relative ease of data collection.[25]

Not only could such a system discover which individual patients would be negatively affected by a given pharmaceutical or intervention, but it could also identify that subpopulation of patients who would benefit from a test or treatment of statistically low efficacy (e.g., a specialized form of chemotherapy). This use of clinical data to individualize tests and treatments is referred to as phenomics and will supplement decision support guided by pharmacogenetics and genomics. It is also possible that there are common interventions for which a positive short-term result is attained but longer term adverse effects manifest themselves at a later date. As such, these outcomes may be otherwise extremely difficult to connect with their true causes. Finally, this system might be used to find novel indications for currently available drugs, providing a fast track to authorization for prescription of currently off-label uses. In this case, the risk profile of the drug would have already been assessed over a long period of time and in many clinical contexts.

A key implementation challenge is the protection of patient privacy. The MIMIC database thoroughly de-identifies all patient records. To fulfill the potential of this proposed open-access database, it is also necessary to link separate data sources. This is hampered by the lack of a unique patient identifier in the United States. Many provider organizations have developed reasonably reliable data linkage algorithms based on patient demographic information. However, increased data linkage renders the process of deidentification more difficult. One possible approach is to formally regulate the practice of data linkage, defining what is legal and ethical, and to include patients in a public forum that frames this policy.[26]

To fully utilize the capacity of this system, it would need to be seamlessly integrated into the clinicians' workflow. The most likely method of integration would be within a patient's EHR, since currently it is the EHR that is populated with information that is actionable and unique to the clinical context of the documented healthcare encounter. A model for integrating this approach into patient care is described in a recent article by clinicians from Stanford University School of Medicine.[27] When treating a young girl with systemic lupus erythematosus and a complex set of comorbidities, the published literature provided insufficient guidance for clinical decision making. However, clinicians were able to use immediate advanced text-searching capabilities to query their institution's electronic medical record data warehouse to review the outcomes of similar patients. This enabled them to adopt a data-supported approach to the treatment decision.

It has heretofore been the usual practice for medical societies to utilize their own expert opinion as the basis for formulation of clinical practice guidelines within their respective fields. However, this process has not always been fully evidence based and, more worrisome, has been plagued by conflicts of intellectual and financial interests.[28] Ultimately, the information system that we are proposing would represent a helpful additional resource for the formulation of objective, robust, and outcome-based clinical guidelines.

An important development that the FDA can leverage in expanding the scope of the Sentinel Initiative is the creation of the National Patient-Centered Clinical Research Network called PCORnet.[22] It consists of 11 clinical data research networks (CDRN) across the country that securely collect health information during the routine course of patient care. In contrast to the Sentinel Initiative's distributed database model, where data partners alone controlled access to their data, data sharing across the network is integral to PCORnet and will be accomplished using a variety of methods that ensure confidentiality by preventing patient re-identification. In addition, PCORnet also includes 18 patient-powered research networks that will be operated and governed by groups of patients who are interested in sharing health information and participating in research. The Scalable Collaborative Infrastructure for a Learning Healthcare System, which was built with open source modular components, will be employed to enable a queryable semantic data model that plugs universally into the point of care.[29] If all goes well, by September 2015, PCORnet will be a giant repository of medical information from 26 to 30 million Americans. The size of one CDRN database is projected to be at least 10 terabytes (Kenneth Mandl, personal communication). With the advent of cloud computing and efficient parallel, distributed statistical algorithms for big data, such as MapReduce and Hadoop,[30] performing computationally intensive analytics on terabytes of heterogeneous patient data records is not just feasible, but scalable. The cost of such a system will be driven primarily by the governance of its use rather than the technology, and should be justifiable if it can

"IF ALL GOES WELL, BY SEPTEMBER 2015, PCORNET WILL BE A GIANT REPOSITORY OF MEDICAL INFORMATION FROM 26 TO 30 MILLION AMERICANS."

indeed support the type of research described in this article—from pharmacovigilance to clinical care optimization.

Conclusions

It is likely that many stakeholders would benefit from the creation of a de-identified and open patient EHR database. For example, this active pharmacovigilance system would assist pharmaceutical companies in their risk profile construction and could be used to collect data on their products within widely diverse clinical contexts, including those involving patients with comorbidities, and who were excluded in phase 3 clinical trials. This richer data warehousing approach could help pharmaceutical companies comply more rapidly and effectively with FDA requirements on postmarket surveillance. In addition, analysis of routinely collected clinical data might yield cost savings compared to the bespoke data collection of RCTs and many longitudinal epidemiological studies.

Because so many institutions, from academia to healthcare to industry, can benefit from the use of the technology we have described, it is logical to advocate implementation of these systems within an open data framework. Active surveillance need not be utilized exclusively by the FDA to make regulatory decisions on whether or not to approve a pharmaceutical, and for what conditions it should be approved. Ultimately, it is the well-informed patients, providers, and payers that can mitigate the harmful public health concerns associated with poorly conceived or unduly influenced policy recommendations. The open data framework would represent a public good by providing the pieces necessary to create viable and effective systems for the creation of more fully evidence-based medicine, as well as an efficient and all-encompassing system of clinical care optimization. This should all be done in a context that allows for the introduction of necessary elements of innovation as well as supporting patterns of standardization. If this information system is successfully planned and executed, including a proper balance of patient and provider privacy with the necessary data accessibility features, it could be a significant step toward broad and seamless quality improvement in healthcare.

Author Disclosure Statement

None of the authors have any financial conflict of interest to disclose.

References

1. Bohmer RM. Designing Care: Aligning the Nature and Management of Health Care. Cambridge: Harvard Business Press, 2009.
2. Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med 2013; 159:560–562.
3. Humphreys K, Maisel NC, Blodgett JC, et al. Extent and reporting of patient nonenrollment in influential randomized clinical trials, 2002 to 2010. JAMA Intern Med 2013; 173:1029–1031.
4. Evans A, Kalra L. Are the results of randomized controlled trials on anticoagulation in patients with atrial fibrillation generalizable to clinical practice? Arch Intern Med 2001; 161:1443–1447.
5. Downing NS, Aminawung JA, Shah ND, et al. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005–2012. JAMA 2014; 311:368–377.
6. Stevens W, Normand C. Optimisation versus certainty: understanding the issue of heterogeneity in economic evaluation. Soc Sci Med 2004; 58:315–320.
7. ClinicalTrials.gov. Clinical Trial Trends, Charts, and Maps. National Institute of Health, Bethesda, MD, 2014. Available online at http://clinicaltrials.gov/ct2/resources/trends (Last accessed July 30, 2014).
8. Grande D, Mitra N, Shah A, et al. Public preferences about secondary uses of electronic health information. JAMA Intern Med 2013; 173:1798–1806.
9. Okun S, McGraw D, Stang P, et al. Making the case for continuous learning from routinely collected data. Discussion Paper, Institute of Medicine, Washington, DC, 2013. Available online at www.iom.edu/makingthecase
10. Celi LA, Mark RG, Stone DJ, et al. "Big Data" in the intensive care unit: closing the data loop. Am J Respir Crit Care Med 2013; 187:1157–1160.
11. Scott D, Lee J, Silva I, et al. Accessing the public MIMIC-II intensive care relational database for clinical research. BMC Med Inform Decis Mak 2013; 13:9.
12. Saeed M, Villarroel M, Reisner AT, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 2011; 39:952–960.
13. Dejam A, Malley BE, Feng M, et al. The effect of age and clinical circumstances on the outcome of red blood cell transfusion in critically ill patients. Crit Care 2014 (In Press).
14. Ghassemi M, Marshall J, Singh N, et al. Leveraging a critical care database: selective serotonin reuptake inhibitor use prior to ICU admission is associated with increased hospital mortality. Chest 2014; 145:745–752.
15. Moses C, Celi LA, Marshall J. Pharmacovigilance: an active surveillance system to proactively identify risks for adverse events. Popul Health Manag 2013; 16:147–149.
16. Weaver J, Willy M, Avigan M. Informatic tools and approaches in postmarketing pharmacovigilance used by FDA. AAPS J 2008; 10:35–41.
17. Psaty BM, Breckenridge AM. Mini-sentinel and regulatory science—big data rendered fit and functional. N Engl J Med 2014; 370:2165.
18. Avorn J. The promise of pharmacoepidemiology in helping clinicians assess drug risk. Circulation 2013; 128:745–748.
19. Lee W. Electronic Medical Record Data Sources Sought Notice. Food and Drug Administration, Rockville, MD, 2014. Available online at https://www.fbo.gov/spg/HHS/FDA/DCASC/FDA-SS-1127349/listing.html (Last accessed July 30, 2014).

20. Maguire J, Dhar V. Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type 2 diabetics: data-driven predictive analytics in healthcare. Health Syst 2013; 2:73–92.
21. McCann E. Mayo Clinic, Optum Team Up. Healthcare IT News, Chicago IL, 2013. Available online at www.healthcareitnews.com/news/mayo-clinic-optum-team (Last accessed July 30, 2014).
22. PCORnet: The National Patient-Centered Clinical Research Network. Patient-Centered Outcomes Research Institute, Washington, DC, 2014. Available online at www.pcori.org/assets/National-Patient-Centered-Clinical-Research-Network-description-FINAL.pdf (Last accessed July 30, 2014).
23. Celi LA, Zimolzak A, Stone DJ. Dynamic Clinical Data Mining: search engine-based decision support. J Med Internet Res 2014; 2:e13.
24. Berwick D, Jain S, Porter M. Clinical registries: the opportunity for the nation. HealthAffairs Blog, May 2011. Available online at http://healthaffairs.org/blog/2011/05/11/clinical-registries-the-opportunity-for-the-nation/ (Last accessed July 30, 2014).
25. Lauer MS, D'Agostino RB. The randomized registry trial—the next disruptive technology in clinical research? N Engl J Med 2013; 369:1579–1581.
26. Weber G, Mandl K, Kohane I. Finding the missing link for big biomedical data. J Am Med Assoc 2014; 311: 2479–2480.
27. Frankovich J, Longhurst CA, Sutherland SM. Evidence-based medicine in the EMR era. N Engl J Med 2011; 365:19.
28. Lenzer J. Why we can't trust clinical guidelines. Br Med J 2013; 346:f3830.
29. Mandl KD, Kohane IS, McFadden D, et al. Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): architecture. J Am Med Inform Assoc 2014; 21:615–620.
30. Dong X, Bahroos N, Sadhu E, et al. Leverage hadoop framework for large scale clinical informatics applications. AMIA Jt Summits Transl Sci Proc 2013; 2013:53.

Address correspondence to:

Leo Anthony Celi
Institute for Medical Engineering and Science
Massachusetts Institute of Technology
77 Massachusetts Avenue, E25-505
Cambridge, MA 02139

E-mail: lceli@mit.edu

Appendix