


ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Decision-Analytic Modeling: Past, Present, and Future

Real-World Evidence, Causal Inference, and Machine Learning

William H. Crown, PhD*

OptumLabs,® Cambridge, MA, USA.

ABSTRACT

The current focus on real world evidence (RWE) is occurring at a time when at least two major trends are converging. First, is the progress made in observational research design and methods over the past decade. Second, the development of numerous large observational healthcare databases around the world is creating repositories of improved data assets to support observational research.

Objective: This paper examines the implications of the improvements in observational methods and research design, as well as the growing availability of real world data for the quality of RWE. These developments have been very positive. On the other hand, unstructured data, such as medical notes, and the sparcity of data created by merging multiple data assets are not easily handled by traditional health services research statistical methods. In response, machine learning methods are gaining increased traction as potential tools for analyzing massive, complex datasets.

Conclusions: Machine learning methods have traditionally been used for classification and prediction, rather than causal inference. The prediction capabilities of machine learning are valuable by themselves. However, using machine learning for causal inference is still evolving. Machine learning can be used for hypothesis generation, followed by the application of traditional causal methods. But relatively recent developments, such as targeted maximum likelihood methods, are directly integrating machine learning with causal inference.

Keywords: big data, causal inference, econometrics, epidemiology, machine learning, real-world evidence, targeted maximum likelihood estimator.

VALUE HEALTH. 2019; 22(5):587–592

Introduction

Currently, there is a high level of interest in real-world evidence (RWE). RWE consists of evidence on patient care and health outcomes that has been developed from real-world data on actual clinical practice.[1] Its uses span clinical decision making, regulatory decision making, coverage decisions by payers and health technology assessment, clinical trials design, assessing burden of illness, evaluating market potential for new products, and much more. Key to all these use cases is whether the quality of the RWE is sufficient to support decision making in the context for which it has been developed.

In this *Policy Perspective*, I summarize developments in research design, the proliferation of very large healthcare research data sets around the world, and the advancement of statistical methodology, which has culminated in the current tension between the causal modeling orientation of health services research, econometrics, and epidemiology methods with the classification and prediction methods of supervised machine learning. This is not intended as a formal review of the literature. Nevertheless, with the 20th anniversary of *Value in Health* as a backdrop, I have attempted to draw the reader's attention to several key developments in observational research that have occurred since the journal was launched and consider where we may be headed.

Interest in RWE Around the Globe

With the passage of the American Recovery and Reinvestment Act (ARRA) in the United States in 2009, a new focus on comparative effectiveness research studies was born.[2] Because of concerns about potential rationing of healthcare, ARRA explicitly forbade the use of cost effectiveness in government programs.[3,4] The resulting focus was on noneconomic outcomes but the domain of interventions was very broad—including treatments, diagnoses, care organization models, disease management programs, and much more.

The ARRA legislation also set in motion major changes to the US healthcare research infrastructure through 2 different pathways: (1) meaningful use provisions that led to dramatic

Conflicts of interest: The author has indicated that there are no conflicts of interest with regard to the content of this article.

* Address correspondence to: William H. Crown, PhD, OptumLabs,® One Main Street, Cambridge, MA 02142, USA. Email: william.crown@optum.com

https://doi.org/10.1016/j.jval.2019.03.001

expansion in the volume of electronic health record (EHR) data[5] and (2) establishment of the Patient-Centered Outcomes Research Institute, which established a major new data infrastructure to conduct clinical research and fundamentally changed how researchers thought about the implications of the patient perspective for research design.

Europe has witnessed a parallel set of activities under the rubric of relative effectiveness that assesses whether an intervention provides more benefit than harm relative to a treatment alternative provided in routine care.[6] In particular, the Innovative Medicines Initiative GetReal Consortium encompasses a set of multinational collaborations across Europe to explore the role of using real-world data in drug development and assessment. A recent review of 6 European health technology assessment (HTA) agencies revealed significant variation in their use of real-world data in decision making.[6] Fundamentally, this lack of alignment among the European HTA organizations is a result of differing opinions about the strength of evidence that can be obtained from observational studies.[6]

Interest in RWE studies has its roots in retrospective database analyses to inform payer decision making, comparative effectiveness analysis to inform clinical decision making, and cost-effectiveness analysis that combines the comparative effectiveness data with cost, utility, and other data to evaluate the relative value associated with alternative treatments.[7] Recently, interest in RWE has also been driven by the desire to accelerate the clinical development process that has been so heavily dependent on evidence from randomized controlled trials (RCTs). In the United States, this has taken the form of regulatory mandates on the US Food and Drug Administration (FDA) embedded in key legislation including the Prescription Drug User Fee Act VI[8] and the 21st Century Cures Act[9] to publish draft guidance for RWE applications by the end of fiscal year 2021. In addition, the 21st Century Cures Act (Section 3022) mandates that FDA propose a framework and enact a program to evaluate RWE to support approval of new indications and to satisfy postapproval requirements.

RWE and Causal Inference

Not all RWE is developed for the purposes of regulatory, clinical, or policy decision making. Although causal modeling is an important component of RWE, RWE is a much broader framework involving the use of observational data for market assessment, clinical trials design, measuring burden of disease, understanding variation in practice, and improving clinical aspects of healthcare delivery, to name a few. For example, disease management programs may seek to improve population health by intervening with patients at risk of adverse health events; this requires the ability to predict risk but not the need for causal inference about the effect of a particular factor on that risk. Nevertheless, assessing whether the disease management program has the intended effect of improving healthcare outcomes *does* require causal inference.

Several leading researchers stress that epidemiology requires asking causal questions and advocate for the use of formal causal frameworks to improve the analytical rigor of observational data analysis.[10,11] Epidemiology is the bedrock for making regulatory and clinical decisions about the effectiveness and safety of therapeutic interventions. Similarly, causal inference is needed to support policy decisions by HTA agencies and payers with regard to coverage and reimbursement. And causal inference is needed to assess all manner of program evaluation questions such as benefit design, care organization models, and effectiveness of disease management programs.

The critical importance of using a causal framework to design studies using observational data is appropriately receiving increasing attention. The use of a causal framework forces the researcher to describe the theoretical model to be estimated—in particular, how the key causal effect to be estimated fits within the broader system. The observed data are then compared with those needed to estimate the causal effect within the context of the broader theoretical model. This forces the researcher to confront data limitations and consider their implications for the ability to estimate the causal effect. An important aspect of the ability to estimate causal effects, familiar to econometricians, is that of identification. The concept of identification refers to whether the available data enable the estimation of the system parameters of interest.

Only at this point do statistical estimation methods enter the picture. There are many choices of methods for estimating a causal effect of interest. Each has its own properties and some will perform better than others for specific questions. Nevertheless, the key point is that the choice of statistical methods is a decidedly second-order consideration when conducting studies with observational data. Design is paramount.

Evolving Statistical Methods and Methods Guidance

Over the past 20 years, *Value in Health* has published a series of International Society for Pharmacoeconomics and Outcomes Research Good Practice Guidelines related to the conduct of high-quality epidemiological and outcomes research using observational databases.[1,12–16] Similarly, the International Society for Pharmacoepidemiology, the FDA, the European Medicines Agency, the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance, and the European Network for Health Technology Assessment have all published guidance documents on good practice for the analysis of observational healthcare data.[17–21] These publications recommend that observational research design incorporate pre- and postmeasurement on both intervention and control groups (quasi-experimental design) and the use of matching or weighting methods to achieve high dimensional balance on observable confounders. For studies of pharmaceutical interventions, best practice with respect to study design generally also includes comparing new initiators, controlling for medication adherence, and avoiding reverse causation, immortal time bias, and adjustment for causal intermediaries.

When Can We Trust RWE From Observational Studies?

Despite common perceptions to the contrary, a 2014 Cochrane review of the literature[22] concluded that observational studies usually generate similar average treatment effects (ATEs) to those reported in clinical trials even when no attempt has been made to mirror the inclusion and exclusion criteria of the RCTs. This conclusion echoed that of several previous studies that had also found that the treatment effects obtained from RCTs and observational studies in corresponding disease areas tend to be very similar.[23,24]

Remarkably, these results comparing the similarity of ATEs from RCTs and observational studies have been obtained without examining the subset of observational studies that attempted to mimic the inclusion and exclusion criteria of the trials and other design features. But the result is that there remains confusion in the literature about the importance of randomization in RCTs versus a plethora of design differences between RCTs and observational studies that can easily be responsible for differences in conclusions.

A classic example of disagreement between RCT and observational study results is provided by the Nurses' Health Study and the Women's Health Initiative—both very large, well-designed studies that were conducted a decade apart. The Nurses' Health Study was an observational study that demonstrated convincing evidence about the protective effect of hormone replacement therapy against the development of cardiovascular disease.[25] Nevertheless, 10 years later the Women's Health Initiative, a large RCT, came to the opposite conclusion.[26] Hernan et al[27] noted that the 2 studies measured cardiovascular risk over different time periods and this was an important contributor to the difference in findings. When cardiovascular risk in the Nurse's Health Study was measured over the same 2-year follow-up period as in the Women's Health Initiative, the results of the 2 studies were very similar. Although, in this instance, the observational design and analysis were able to replicate the RCT, the researchers had the RCT results available as a target. This ultimately led to identifying the design problem that had led to the divergent results. Generally, RCT results are not available for comparison, which raises concern that researchers might miss things in observational research design and be led astray. This has led to the publication of several recent articles outlining design principles for improving the reliability of pharmacoepidemiology and economics studies.[11,28,29]

Despite the results cited earlier demonstrating the high degree of agreement generally found between retrospective database studies and RCTs, there is also evidence for wide variation in results from observational studies with similar research designs conducted on different data sets. In a study of 53 drug/outcome pairs using 2 common research designs, Madigan et al[30] estimate that 20% to 40% of database studies can switch from a statistically significant result in one direction to a statistically significant result in the other direction, depending on the data set used for the analysis—even after controlling for common study design.

One possible reason for variation across observational studies is that the patient populations represented in the various samples are very different. But study results can also vary because of the statistical methods used as well as differences in study design. Epidemiological research on drug safety over the past decade has established generally accepted research designs and methodologies for estimating the effects of drug treatment on adverse events and other safety outcomes with observational data.[11,28,29] Using similar methodological techniques that include high dimensional control for confounders using propensity score matching, inverse probability weighting, and related methods, there is reason to believe that well-designed studies using RWE may play an expanded role in regulatory submission for review and approval with respect to additional patient populations and/or new indications for previously approved products. Finally, understanding the potential sources of variation in study results hinges critically on transparency of reporting.[31]

Data Linkage and Causal Inference

There is a worldwide explosion in the volume and types of data available to support outcomes research, health economics, and epidemiology. In the context of the causal modeling framework considered earlier, the growth in data availability offers promise for improved causal inference. Estimates from technology companies generally state that 80% of "big data" is unstructured; this is also the case for biomedical data such as physician narratives in electronic medical records (EMRs) and data streams from medical device monitors.[32] Structured medical data such as claims are also growing rapidly but, although structured, need considerable cleaning and mapping before they are suitable for research purposes. Numerous initiatives are amassing huge repositories of claims and EMR data that have been curated with an eye toward facilitating healthcare research (Health Care Cost Institute, Observational Health Data Sciences and Informatics Network, FDA Sentinel Network, the Patient-Centered Clinical Research Network, OptumLabs® Data Warehouse, Health Data Research UK, Clinical Practice Research Datalink, UK-CRIS Network, the national registries of Sweden, Denmark, Finland, Iceland, and Norway, Taiwan's National Health Insurance Research Database, and many others).[33–42]

One important implication of this expanded data access is the potential of reducing missing variable bias through data linkage. For example, linking claims data with EMRs enables control for confounders that are missing in analyses based on claims data alone. Claims data are generally quite good for capturing the breadth of the experience of patients, their medical comorbidities, the drugs they take, their visits, and so forth, but they are not very good for measuring disease severity, biomarkers, and other clinical factors. In contrast, EMR data are much stronger for capturing clinical detail but they can often be confined to particular treatment settings such as hospitals and oncology clinics. In comparison with claims data, much of the knowledge about comorbidities may be missing. As a result, models estimated with EMR data alone are likely to be biased because they lack important information on comorbidities. Conversely, models estimated with claims data alone are likely to be biased because they lack important controls for clinical severity. The linkage of data sets should help address many of these missing data issues.

The Rise of Machine Learning in Healthcare

Several aspects about the changing healthcare data landscape—the rapid growth in the volume of healthcare data, the fact that much of it is unstructured, the ability to link different types of data together (claims, EHR, sociodemographic characteristics, genomics), the speed with which data are being refreshed—create serious challenges for traditional statistical methods from epidemiology, econometrics, and health services research. As a result, there is growing interest in the use of machine learning to help address these analytic challenges.

The term *machine learning* refers to a large family of mathematical and statistical methods that have historically been focused on classification and prediction.[43] It is beyond the scope of this *Policy Perspective* to describe machine learning methods in detail, but Figure 1 provides a high-level overview of the range of methods. At the highest level, there are 2 broad categories of methods—supervised and unsupervised. Unsupervised methods are focused mainly on dimension reduction and learning the underlying structure of the data. Supervised methods require the specification of an outcome variable and are focused on prediction or classification.[43,44]

Machine learning methods can be particularly powerful tools for satisfying the evidence needs of the broader RWE objectives beyond causal inference alone. The ability to improve predictions of whether someone is at risk of developing a disease or having a health event such as a heart attack is extremely important in healthcare delivery—potentially enabling intervention before adverse outcomes occur. In some areas of medicine, such as radiology, machine learning methods show great promise for improved diagnostic accuracy.[45,46]

The applications of machine learning in healthcare have changed significantly over the last decade. Earlier work often used machine learning methods such as Lasso, random forest, and support vector machines to predict hospitalization[47,48] or the onset of

Figure 1. Types of machine learning methods.

- Unsupervised Learning—Focus on reducing data dimensionality and learning the underlying structure of the data
  - Dimensionality Reduction
    - Principal Components Analysis
    - Latent Variables and Factor Analysis
    - Multidimensional Scaling Methods
  - Clustering
    - K-means Clustering
    - Gaussian Mixture Models
    - Agglomerative Hierarchical Clustering
- Supervised Learning—Focus on classification and prediction.
  - Classification and Prediction
    - Lasso, Ridge, and Elastic Net Regression
    - Random Forest
    - Boosting and Bagging Models
    - Support Vector Machines
    - Neural Networks
    - Learning Ensemble Methods

disease.[49] These models often represented only modest improvements over traditional regression methods. More recently, with the development of graphical processing units and massive volumes of healthcare data, deep learning models have been increasingly used.[50,51] Traditional machine learning models rely on the development of features or variables that are defined on the basis of researcher domain knowledge. In contrast, deep learning models extract features directly from the data—enabling the identification of correlations in the data that may have been previously unknown. A recent study used EHR data from several academic medical centers to predict in-hospital mortality, unplanned readmission, prolonged length of stay, and all of the patients' final discharge diagnoses. These models outperformed traditional clinical predictive models in all cases.[51] Nevertheless, there are basic questions about the readiness of EHR systems to support machine learning methods from a data quality standpoint.[52]

Is Machine Learning Compatible With Causal Inference?

Some machine learning approaches use regression-based methods for prediction. For example, Lasso, Ridge, and elastic net methods use correction factors to reduce the risk of overfitting.[44,53] Moreover, the K-fold cross-validation approach used in machine learning can be thought of as a more sophisticated and systematic version of the best practice of splitting one's sample into 2—one for model development and the other for final model estimation.

Unfortunately, there is nothing magical about machine learning that protects against the usual challenges encountered in observational data analysis. In particular, just because machine learning methods are operating on big data does not protect against bias. Increasing sample size—for example, getting more and more claims data—does not correct the problem of bias if the data set is lacking in key clinical severity measures such as cancer stage in a model of breast cancer outcomes.[54] This is a specific example of the issue of identification mentioned earlier. Machine learning can provide a statistical method for estimating causal effects but only in the context of an appropriate causal framework. The use of machine learning without a causal framework is fraught with danger.

Nevertheless, economists, epidemiologists, and health services researchers have been trained that they must have a theory that they test through model estimation. The major limitation of this approach is that it makes it very difficult to escape from the

confines of what we already know (or think we know). Traditional machine learning is agnostic to theory and lets the data speak for itself. This makes supervised machine learning very powerful for exploratory analysis. One promising approach might be to use the power of machine learning to identify all the features in a data set that are correlated with an outcome of interest. These features then become candidates for inclusion in a causal modeling framework that is then estimated using standard methods from economics or epidemiology.[55]

A second approach is to use machine learning methods to estimate the causal treatment effects directly. An example of this approach is provided by targeted maximum likelihood estimation (TMLE).[56,57] TMLE uses doubly robust maximum likelihood estimation to update the initial model using estimated probabilities of exposure.[58] The ATE is then estimated as the average difference in the predicted outcome for treated patients versus their outcome if they had not been treated (their counterfactual). Because the ATE is based on the predicted values, rather than an estimated parameter value, the modeling methods can draw upon the full repertoire of machine learning and traditional econometric/epidemiological methods. TMLE appears to outperform traditional methods such as propensity score matching and inverse probability weights in simulation.[56] Finally, using ensemble methods produces estimates that are asymptotically as good as the best-performing model—eliminating the need to make strong assumptions about functional form and estimation method up front.[56,57]

Conclusions

We are witnessing the convergence of several major trends: (1) a focus on RWE and its potential role in generating high-quality evidence previously reserved for RCTs; (2) dramatically expanding data—some of it in curated research data assets—but most in need of a great deal of work before being suitable for research; and (3) the arrival of machine learning methods.

Although RWE is a relatively new term, retrospective database studies have been used for clinical trials design and generating evidence on disease burden, comparative/relative effectiveness, and safety surveillance for at least 2 decades. Such studies were mainly conducted with claims data and many fell short of the level of rigor that we would now consider necessary to draw causal inferences from observational data. Over the years, however, observational research designs and statistical methods have gotten much stronger. Most recently, the arrival of machine learning methods is improving the ability to identify patients at risk of healthcare events so that providers can intervene to prevent adverse health outcomes. Predictive analytics are rapidly being upgraded using machine learning methods to take advantage of this opportunity. But we are still left with the question of drawing causal inference from observational data. Can machine learning help? At a minimum, it seems that machine learning would be an effective tool for hypothesis generation. A core strength of machine learning is to identify correlational structures in observational data. Once identified, these structures can be tested with traditional causal modeling approaches. The current frontier is using machine learning to estimate causal models directly. These methods, such as TMLE, use machine learning as a statistical estimator but do so within a causal modeling framework. More evidence is needed, but whether machine learning methods are used for prediction alone, or for causal inference, it seems clear that they are here to stay.

Source of Financial Support

The author has no other financial relationships to disclose.

REFERENCES

1. Berger M, Sox H, Willke R, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value Health*. 2017;20(8):1003–1008.
2. US Congress. American Recovery and Reinvestment Act of 2009. PL 111-5, 111th Congress of the United States. 2009.
3. Chandra A, Jena A, Skinner J. The pragmatist's guide to comparative effectiveness research. *J Econ Perspect*. 2011;25(2):27–46.
4. Garber A, Sox H. The role of costs in comparative effectiveness research. *Health Aff (Millwood)*. 2010;29(10):1805–1811.
5. Evans R. Electronic health records: then, now, and in the future. *Yearb Med Inform*. 2016;(suppl 1):S48–S61.
6. Makady A, ten Ham R, de Boer A, et al. Policies for use of real-world data in health technology assessment (HTA): a comparative study of six HTA agencies. *Value Health*. 2017;20(4):520–532.
7. Malone D, Brown M, Hurwitz J, Peters L, Graff J. Real-world evidence: useful in the real world of US payer decision making? How? When? And what studies? *Value Health*. 2018;21(3):326–333.
8. H.R.34-21st Century Cures Act. Public Law No. 114-255. https://www.congress.gov/bill/114th-congress/house-bill/34/text. Accessed April 3, 2019.
9. Prescription Drug User Fee Act (PDUFA). https://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm446608.htm. Accessed April 3, 2019.
10. Petersen M, van der Laan M. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology*. 2014;25(3):418–426.
11. Goodman S, Schneeweiss S, Baiocchi M. Using design thinking to differentiate useful from misleading evidence in observational research. *JAMA*. 2017;317(7):705–707.
12. Motheral B, Brooks J, Clark MA, et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Value Health*. 2003;6(2):90–97.
13. Berger M, Mamdani M, Atkins D, Johnson M. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—part I. *Value Health*. 2009;12(8):1044–1052.
14. Cox E, Martin B, Van Staa T, Garbe E, Siebert U, Johnson M. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—part II. *Value Health*. 2009;12(8):1053–1061.
15. Johnson M, Crown W, Martin B, Dormuth C, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Practices for Retrospective Database Analysis Task Force Report—part III. *Value Health*. 2009;12(8):1062–1073.
16. Berger M, Martin B, Husereau D, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force Report. *Value Health*. 2014;17(2):143–156.
17. Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf*. 2016;25(1):2–10.
18. FDA. Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf; 2013. Accessed April 3, 2019.
19. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) module VIII—post-authorization safety studies (rev. 2). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf. Accessed January 4, 2019.
20. EUTneHA. http://www.eunethta.eu/outputs/Internal-Validity-of-non-randomized studies-(NRS)-on-interventions. Accessed January 4, 2019.
21. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. The European Union Electronic Register of Post-Authorization Studies (EU PAS Register). http://www.encepp.eu/encepp_studies/indexRegister.shtml. Accessed January 4, 2019.
22. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014;4:MR000034.
23. Concato J, Shah N, Horwitz RI. Randomized controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887–1892.
24. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342(25):1878–1886.

25. Stampfer MJ, Colditz GA, Willett WC, et al. Postmenopausal estrogen therapy and cardiovascular disease: ten-year follow-up from the Nurses' Health Study. *N Engl J Med*. 1991;325(11):756–762.
26. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus projestin in healthy post-menopausal women. *JAMA*. 2002;288(3):321–333.
27. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779.
28. Franklin J, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther*. 2017;102(6):924–933.
29. Fralick M, Kesselheim A, Avorn J, Schneeweiss S. Use of health care databases to support supplemental indications of approved medications. *JAMA Intern Med*. 2018;178(1):55–63.
30. Madigan D, Ryan P, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol*. 2013;178(4):645–651.
31. Wang S, Schneeweiss S, Berger M, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Value Health*. 2017;20:1009–1022.
32. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. *Yearb Med Inform*. 2014;9:14–20.
33. Health Care Cost Institute. https://www.healthcostinstitute.org. Accessed January 4, 2019.
34. Observational Health Data Sciences and Informatics. https://www.ohdsi.org. Accessed January 4, 2019.
35. National Patient-Centered Clinical Research Network. https://www.pcornet.org. Accessed January 4, 2019.
36. Curtis L, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)*. 2015;33(7):1178–1186.
37. Wallace P, Shah N, Dennon T, Bleicher P, Crown W. Optum Labs: building a novel node in the learning health care system. *Health Aff (Millwood)*. 2014;33(7):1187–1194.
38. Health Data Research UK. https://www.hdruk.ac.uk. Accessed February 18, 2019.
39. Clinical Practice Research Datalink. https://www.cprd.com. Accessed January 4, 2019.
40. Clinical Record Interactive Search System (CRIS) Network. https://crisnetwork.co. Accessed January 4, 2019.
41. Maret-Ouda J, Tao W, Wahlin K, et al. Nordic registry-based cohort studies: possibilities and pitfalls when combining Nordic registry data. *Scand J Public Health*. 2017;45(17 suppl):14–19.
42. Hsiao F, Yang C, Huang Y, Huang W. Using Taiwan's national health insurance research databases for pharmacoepidemiology research. *J Food Drug Anal*. 2007;15(2):99–108.
43. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd ed. New York, NY: Springer-Verlag; 2009.
44. Machine learning for health outcomes research. *Value Health*. 2019. In press.
45. Obermeyer Z, Emanuel E. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216–1219.
46. Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–2223.
47. Hong WS, Haimovich AD, Taylor RA. Predicting hospitalization admission at emergency department triage using machine learning. *PLoS One*. 2018;13(7):e0201016.
48. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital admissions. *J Biomed Inform*. 2015;56:229–238.
49. Yu W, Liu T, Valdez R, et al. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Making*. 2010;10:16.
50. Shickel B, Tighe P, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2018;22(5):1589–1604.
51. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18.
52. Perry W, Hossain R, Taylor R. Assessment of the feasibility of automated, real-time clinical decision support in the emergency department using electronic health record data. *BMC Emerg Med*. 2018;18(1):19.
53. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58:267–288.
54. Crown W. Potential application of machine learning in health outcomes research—some statistical cautions. *Value Health*. 2015;18(2):137–140.
55. Rubin DB. Estimating causal effects of treatments in randomized and non-randomized studies. *J Educ Psychol*. 1974;66(5):688–701.
56. Schuler M, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol*. 2017;185(1):65–73.
57. Van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, NY: Springer-Verlag; 2011.
58. Funk MJ, Westreich D, Wiesen C, Sturmer T, Brookhart A, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol*. 2011;173(7):761–767.