Real-World Data, Machine Learning and Causal Inference

Yaru Shi, PhD, and Jie Chen, PhD
Biostatistics and Research Decision Sciences (BARDS)
Merck & Co., Inc., Kenilworth, NJ

The ASA Biop Regulatory-Industry Statistics Workshop
Washington Marriott Wardman Park

In memory of late Joseph F. Heyse who was a mentor, a role model and a good friend of mine and many others.

September 23–25, 2019

Acknowledgements

- Keaven Anderson, BARDS, Merck & Co., Inc., Kenilworth, NJ
- Ray Bain, BARDS, Merck & Co., Inc., Kenilworth, NJ
- Zhen Li, Department of Statistics, North Carolina State University
- Richard Baumgartner, Merck & Co., Inc., Kenilworth, NJ
- Jinghua He, Merck & Co., Inc., Kenilworth, NJ
- Shahrul Mt-Isa, Merck Sharp & Dohme, London

INVENTING FOR LIFE

Objectives

What will and will not be covered in this talk

- Topics to be covered
  - An overview of real-world data (RWD)
  - An overview of machine learning (ML) methodologies
  - An outline of ML in causal inference using RWD
  - A case study in optimal treatment rules
- Topics not to be covered
  - Detail description of RWD
  - Technical details of ML methodologies
  - Detail description of causal inference methodologies
  - More complex applications of ML methods in causal inference (e.g., DTR)
  - ML and causal inference in the development of precision medicine

INVENTING FOR LIFE

Outline

1. Real World Data and Evidence

Introduction

The 21st Century Cures Act

Real-world data and evidence have been increasingly used in regulatory and healthcare decision-making since the passage of the 21st Century Cures Act on December 9, 2016.

- The United States Food and Drug Administration (FDA) has developed a framework for evaluating real-world data and guidance to industry on real-world evidence
  - to support approval of new drugs or new indications for previously approved drugs, or medical devices, and
  - to support post-approval studies for monitoring safety and effectiveness.
- Biopharmaceutical companies use real-world data and evidence
  - to guide clinical trial design and
  - to provide supplementary information for regulatory approval of new product or indication

Introduction

The 21st Century Cures Act

Real-world data and evidence have been increasingly used in regulatory and healthcare decision-making since the passage of the 21st Century Cures Act on December 9, 2016.

- The healthcare community uses real-world data and evidence to
  - to develop guidelines for decision-making to support medical practice and
  - to assess treatment patterns, costs and outcomes of interventions.
- High-performance computing and machine learning algorithms are available and have been conveniently applied to the analysis of real-world data
- There are still substantial challenges in deriving real-world evidence from real-world data and in using the evidence for regulatory and healthcare decision-making

What Are Real-World Data

The foundation of RWE generation, but it is not just a single entity

Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (FDA, 2018).

- There are various types of RWD that can provide a valuable foundation to research alone or as part of a broader evidence platform.
- Understanding different types of RWD and their attributes is critical in selecting appropriate RWD for decision-making purposes.
  - Electronic medical records may be more appropriate for studying disease natural history
  - Health insurance claims data may be appropriate for cost-effectiveness studies
  - Pharmacy dispensing data may be appropriate for studying patient's medication adherence
- Understanding the merits and limitations of each is important to select fit-for-purpose RWD.

INVENTING FOR LIFE

What Are Real World Data (1/5)

Commonly used RWD types

Examples of common RWD sources (Makady et al., 2017)

1. Electronic medical/health records (EMR/EHR): Digital patient records comprising structured and unstructured data fields including patient's demographics, clinical narratives, diagnosis, procedures and medications
2. Health insurance claims data: Data with structured fields on healthcare utilization, prescription, medical claims, and health plan information.
3. Pragmatic clinical trials (PCTs): PCTs are designed to show the real-world effectiveness of an intervention in broad patient groups and therefore to inform a clinical or policy decision by providing evidence for adoption of the intervention into real-world clinical practice (Ford and Norrie, 2016).

What Are Real World Data (2/5)

Commonly used RWD types

4. Disease registries: Data on patient with a particular disease (usually chronic disease), such as breast cancer registry, SEER (Surveillance, epidemiology, and endpoint results) registry (See https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries for a complete list of NIH maintained registries).
5. Product registries: Databases that contain information on product usage, concomitant medications, safety, QoL, and product satisfaction and that provide manufacturers and regulatory agencies with information on product performance and safety outside of the RCT.
6. Pharmacy dispensing databases: Data containing prescription dispensing and refill, costs and payment information.

What Are Real World Data (3/5)

Commonly used RWD types

7. Epidemiological observational databases: Epidemiological studies on selected populations (e.g., children, elderly, pregnant women) or communities (e.g., counties, states, or countries) for surveillance of events. Examples are OMOP (Observational Medical Outcomes Partnership) and OHDSI (Observational Health Data Sciences and Informatics).
8. Omics related databases: Databases capturing information on patient physiology, biology, health, behavior and possible environmental interaction. Examples are datasets on pharmacogenomics, metabolomics, and proteomics.
9. Mortality databases: Databases containing human mortality information, such as Berkeley Mortality Database and WHO mortality database.

What Are Real World Data (4/5)

Commonly used RWD types

10. Community-based databases: Databases recording longitudinal health status, medical practice, social-health behaviors of the population in a selected community, such as Framingham Study and Clinical Practice Research Datalink (CPRD)
11. Public health databases: Databases containing public health related information, such as PubMed, PLOS, and Cochrane Open Access.
12. Data generated from mobile devices and wearables: Databases with information on subject's daily activities, physiological indicators, ECG, and brain activities. Examples are data generated from smart phones, smart watches and biosensors (Bagot et al., 2018).
13. Social media: Patient's social media posts that contain product usage, self-reported outcomes, and other healthcare related information. Examples of social media data are Twitter and Live Tweet.

What Are Real World Data (5/5)

Commonly used RWD types

14. Other special data sources: Databases created for special purposes such as National Sentinel Program, FAERS (FDA Adverse Event Reporting System), VAERS (Vaccine Adverse Event Reporting System), VSD (Vaccine Safety Datalink), and National Immunization Programs.

Real World Data and Evidence and AI

Increasing use of artificial intelligence

Fueled by the available computational tools in both hardware and software and the accelerating use of big data (including human genomes and real-world data), artificial intelligence (AI) plays an increasingly important role in biopharmaceutical and healthcare industry.

- Davenport and Ronanki (2018) break down AI techniques into three types of applications:
  1. Process automation - automation of digital and physical tasks
  2. Cognitive insight - algorithms to detect patterns in vast volumes of data and interpret their meaning
  3. Cognitive engagement - natural language process and intelligent agents customer service

- The use of AI in pharmaceutical industry is primarily in the first two categories – compound screening and profiling, patient feature selection, and dynamic treatment adaptation for developing precision medicines, disease diagnosis, health technology assessment, outcome prediction, public health surveillance, and real-world trials for delivering precision healthcare to the needed patients.

INVENTING FOR LIFE

Real World Data and Evidence and AI (1/3)

AI and machine learning in biomedical field

A few examples of successful applications of AI and machine learning methodologies in biopharmaceutical and healthcare industry include:

- Research scientists deploying AI to discover lead drug compounds (Fleming, 2018; Hessler and Baringhaus, 2018; Vogt et al., 2018).
- Cardiologists applying AI techniques in cardiovascular medicine to explore novel genotypes and phenotypes to improve the quality of patient care, enable cost-effectiveness, and reduce readmission and mortality rates (Krittanawong et al., 2017).
- Oncologists using AI for tumour segmentation, histopathological diagnosis, tracking tumour development, and prognosis prediction (Londhe and Bhasin, 2018).

Real World Data and Evidence and AI (2/3)

AI and machine learning in biomedical field

- Ophthalmologists applying machine learning (ML) and particularly deep learning (DL) methods to identify, localize and quantify pathological features in almost every macular and retinal disease (Schmidt-Erfurth et al., 2018).
- Radiologists use AI and machine learning methodologies to analyze magnetic resonance imaging and computed tomography for diagnosis of diseases and their progression (Shen et al., 2017; Pesapane et al., 2018).
- Data scientists and clinicians apply predictive modeling with electronic health record (EHR) data to drive personalized medicine and improve healthcare quality (Obermeyer and Emanuel, 2016; Rajkomar et al., 2018).
- There are numerous examples showing vast opportunities and possibilities of AI and machines learning applications in medicine and healthcare

Real World Data and Evidence and AI (3/3)

AI and machine learning in biomedical field

- cutting-edge AI and machine learning methodologies and their applications to precision medicine and healthcare in particular

Big and High-Dimensional Data (1/2)

Big and high-dimensional data poses numerous challenges

- Big data, often high-dimensional, are collected in structural or nonstructural manner with a variety of purposes that may not be research oriented
- The more the better? In general, no. The more the high-quality data the better.
- Understand the data (Shiffrina, 2016)
  - Why are the data collected, for what purposes, from which population?
  - Are there potentially interesting patterns of association in the data? How to find and interpret them? How to deal to a large number of mostly uncontrolled confounders and covariates (some may be unmeasured))with correlations among them, and between them and the identified variables
  - How to define causality and the degree of causality in the big data?
  - Can the patterns of association, if exist, be interpreted as causality? Why or why not?

Big and High-Dimensional Data (2/2)

Big and high-dimensional data poses numerous challenges

- Big and high-dimensional data can help in causal inference (Monroe et al., 2015)
  - Better design future experiments
  - Leverage diversity or heterogeneity with precise subpopulation
  - Reveal patterns that previously were difficult to observe
- How to analyze the massive data? (Grimmer, 2015)
  - "...combining machine learning to make causal inferences is one of the fastest growing and most open fields"
- Traditional causal inference relies on DAGs that may not be available for big data
- Data-driven learning: Regression, prediction and classification
- Causal structural learning methodologies: Classical greedy search over DAGs, reinforcement learning

Causal Inference Using Real-World Data

Causal roadmap

Balzer et al. (2016) outline a seven-step general framework for causal inference

1. Define the scientific question,
2. Articulate the causal knowledge and the limits of that knowledge with respect to the question,
3. Specify the structural causal models (SCM) including causal diagrams and structural equation models,
4. Identify suitable observed data and their connection with the SCM,
5. Assess identifiability of causal parameters as some function of the observed data,
6. Estimate the corresponding causal parameter using appropriate methods,
7. Interpret the causal results.

2. Machine Learning and Causal Inference

Machine Learning Methodologies

Tree-based methods and super-learning (Chen et al., 2018; Kreif and DiazOrdaz, 2019)

- Classification and regression trees (CART): Construction of prediction models through recursive partitioning of training datasets in order to obtain subsets that are as homogeneous as possible in a given target class
- Random forest: Using bootstrapped samples of the data to grow a tree with a random subset of covariates to create the splits (and thus the leaves)
- Bayesian additive regression trees (BART): Using a likelihood for data and a set of priors for the structure and leaf parameters, in which the priors provide regularisation that prevents any single regression tree from dominating the total fit
- Boosting: An ensemble method for improving the model predictions of any given learning algorithm (e.g., xgBoosting)
- Super learning ensemble: Combining results from several base models to create one that outperforms single models

Machine Learning and Causal Inference

How ML can help in big data and causal inference

- Causality discovery: detection of causal relationship of an exposure with an outcome from data
- Causal effect estimation: incorporating supervised ML tools into estimators for causal parameters such as the average treatment effect under the exchangeability (unconfoundedness) and positivity (overlap) assumptions
- Prediction: e.g., predicting exposure using instrumental variables which can then be used to predict outcome, predicting counterfactuals with big data
- Pattern recognition through data mining and reinforcement learning to determine heterogeneous treatment effect and assess the robustness of estimates to model selection

Learning Causal Relationship – Causal Discovery (1/2)

ML for causal discovery

Examining whether a causal relationship exists

- Constraint-based algorithm: Learning a set of causal graphs that satisfy the conditional independence based on *faithfulness* assumption
  - Peter-Clark algorithm (Spirtes et al., 2000): Working in two-step – Learns undirected graphs and then detects the directions of the edges to return an equivalent class of causal graphs
  - Inductive causation (IC) algorithm: Through an efficient graphical criterion for deciding equivalence of causal models and providing a theoretical basis for extracting causal structures from empirical data (Pearl, 2009).

- Score-based algorithm: Replacing conditional independence tests with the goodness of fit tests by learning causal graphs by maximizing the scoring criterion which returns the score of the causal graph given the data, e.g., Bayesian information criterion (Schwarz et al., 1978), greedy equivalence search (Chickering, 2002) and its extension (Ramsey et al., 2017).

Learning Causal Relationship – Causal Discovery (2/2)

ML for causal discovery

- Functional causal models: Expressing a covariate as a function of its directed causes, which enables to differentiate between different DAGs from the same equivalent class, e.g., the linear non-Gaussian acyclic model (Shimizu et al., 2006, 2011)

Learning Causal Effects (1/2)

ML for learning causal effects

- Without unmeasured confounding – All confounders are measured (Guo et al., 2019)
  - Multiple regression
  - Propensity score matching, stratification and subclassification, IPTW
  - Doubly robust method (DRM) (Funk et al., 2011): Combining multiple regression of outcome and PS method for modeling covariates
  - TMLE (Target maximum likelihood estimation): One of DRMs aiming at minimizing the bias and variance of targeted causal parameters by sacrificing the biases and variances of nuisance parameters (van der Laan and Rose, 2018)
  - Approximate residual balancing (Athey et al., 2018): Bias reduction in learning average treatment effect by combining balance weights with a regularized regression adjustment in high-dimensional data
- With unmeasured confounding

Learning Causal Effects (2/2)

ML for learning causal effects

- Instrumental variable (IV): A valid IV causally influences the outcome only through affecting the treatment, e.g., linear structural causal models for IV estimator, an IV estimator under counterfactual framework
- Front-door criterion (Pearl et al., 2016): A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if (1). Z intercepts all directed paths from X to Y. (2). There is no unblocked path from X to Z. (3). All back-door paths from Z to Y are blocked by X.
  - If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by

$$P(y|do(x)) = \sum_z P(z|x) \sum_x P(y|x', z)P(x') \tag{1}$$

- Regression discontinuity: Using treatment assignment based on some "running variables" which may partially be observable

Machine Learning and Causal Inference

Variable selection (Kreif and DiazOrdaz, 2019)

- A challenging issue in causal inference is the control for confounding, which requires the selection of adjustment set of covariates to make conditional exchangeability hold
- In principle, domain knowledge should be used to select the minimally sufficient set of covariates (Rubin, 2007; Pearl, 2009).
- Back-door criteria to identify adjustment set of covariates (possible confounders) (Pearl et al., 2016).
- In practice, prior knowledge may not be available or may be incomplete regarding confounding variables.
- Data-adaptive approaches to selection of confounding covariates – Belloni et al. (2014) propose a variable selection method that takes into account both covariate-treatment assignment and covariate-outcome association (hence referred to as "double selection").

INVENTING FOR LIFE

Machine Learning and Causal Inference (1/3)

Reinforcement learning

Learning for sequential decision-making, which involves action A, state Z and reward Y. For Markov decision process, the current state z_t and action a_t together determine the next state z_{t+1}, and the reward at the next state y_{t+1} depends on (z_{t+1}, a_t).

- Q-learning ("quality" learning) is based on regression models of outcome on patient covariates at each decision point and is implemented via backward recursive fitting algorithm for deducing optimal dynamic decisions (Nahum-Shani et al., 2012; Schulte et al., 2014).
- A-learning ("advantage" learning) uses the same recursive strategy for outcome regression models that contrast among treatments and for the probability of observed treatment assignment given patient covariates at each decision point (Murphy, 2003; Robins et al., 2004).

Machine Learning and Causal Inference (2/3)

Reinforcement learning

- C-learning ("classification" learning) involves optimization to sequentially minimize a weighted expected misclassification error (Zhang and Zhang, 2018)
- SARSA (State-Action-Reward-State-Action) is an on-policy temporal-difference control method that chooses the action for each state during learning by following a certain policy in order to estimate $Q_\pi(s, a)$ for the current policy Π and all state-action $(s - a)$ pairs (Sutton and Barto, 2018)
- Deep Q network combines Q-Learning and Deep Learning to yield Deep Q Networks, which is suitable for a large number of states and actions (or decision points) (Wang et al., 2015)
- Deep Deterministic Policy Gradient (DDPG) is an algorithm uses off-policy data and the Bellman equation to learn the Q-function, and uses the Q-function to learn the policy (Silver et al., 2014).

Machine Learning and Causal Inference (3/3)

Reinforcement learning

- Blip function (van der Laan and Luedtke, 2015): Based on structural nested mean models of Murphy (2003) and Robins (2004) that assume a parametric model for the "blip function" defined as the additive effect of a blip in current treatment on a counterfactual outcome, conditional on the observed past, in the counterfactual world in which future treatment is assigned optimally.
- Inverse probability of censoring weighted (IPCW): Given a weight to a subject that is inversely proportional to an estimate of the conditional probability of having remained uncensored until time t (Robins and Finkelstein, 2000).
- Inverse probability of treatment weighting (IPTW): A standardized technique that reweights each observation in the sample by taking the reciprocal (hence inverse) of the probability of receiving treatment in the target population (Robins et al., 2000).

3. A Case Study – Optimal Treatment Rules

A Case Study – Optimal Treatment Rules (1/3)

HIV trial with CD4 count as a primary endpoint

A clinical trial compared monotherapy with zidovudine or didanosine, with combination therapy with zidovudine and didanosine, or with zidovudine and zalcitabine, in adults infected with the human immunodeficiency virus type 1 (HIV-1) whose CD4 cell counts were from 200 to 500 per cubic millimeter (Hammer et al., 1996).

- Consider two treatments: zidovudine + didanosine ($A = 1$) versus zidovudine ($A = 0$)
- 2139 HIV-infected patients with either $A = 1$ or $A = 0$.
- Select 532 HIV-infected patients who took zidovudine before the study
- The primary endpoint was CD4 count at 96 ± 5 weeks.
- Consider two covariates: (a) patient's weight (S1) (b) baseline CD4 T cell count (S2)
- Empirical CD4 count at 96 ± 5 weeks: 383.182.

A Case Study – Optimal Treatment Rules (2/3)

HIV trial with CD4 count as a primary endpoint

- Reinforcement learning methods are applied to (a) obtain optimal treatment options (d^{opt}) and (b) estimate CD4 count at optimal treatment d^{opt} (Table 1)

Table 1: Optimal treatment d^{opt} and estimated CD4 count under d^{opt}

Method	d^{opt}	$\widehat{E_{d^{\text{opt}}}Y}$
QMR	$I(13.60 + 1.19S_1 - 0.26S_2 > 0)$	397.162
QRF	*	395.090
A-Learning	$I(13.60 + 1.19S_1 - 0.26S_2 > 0)$	397.162
C-Learning (TMLE)	$I(-41.66 + 0.96S_1 - 0.10S_2 > 0)$	400.764
C-Learning (IPTW)	$I(-41.66 + 0.96S_1 - 0.10S_2 > 0)$	399.540
Blip (TMLE)	*	396.615
Blip (IPTW)	*	392.619
DR–IPCW (TMLE)	*	395.847
DR–IPCW (IPTW)	*	391.824

* denotes optimal treatment rule without explicit form.

A Case Study – Optimal Treatment Rules (3/3)

HIV trial with CD4 count as a primary endpoint

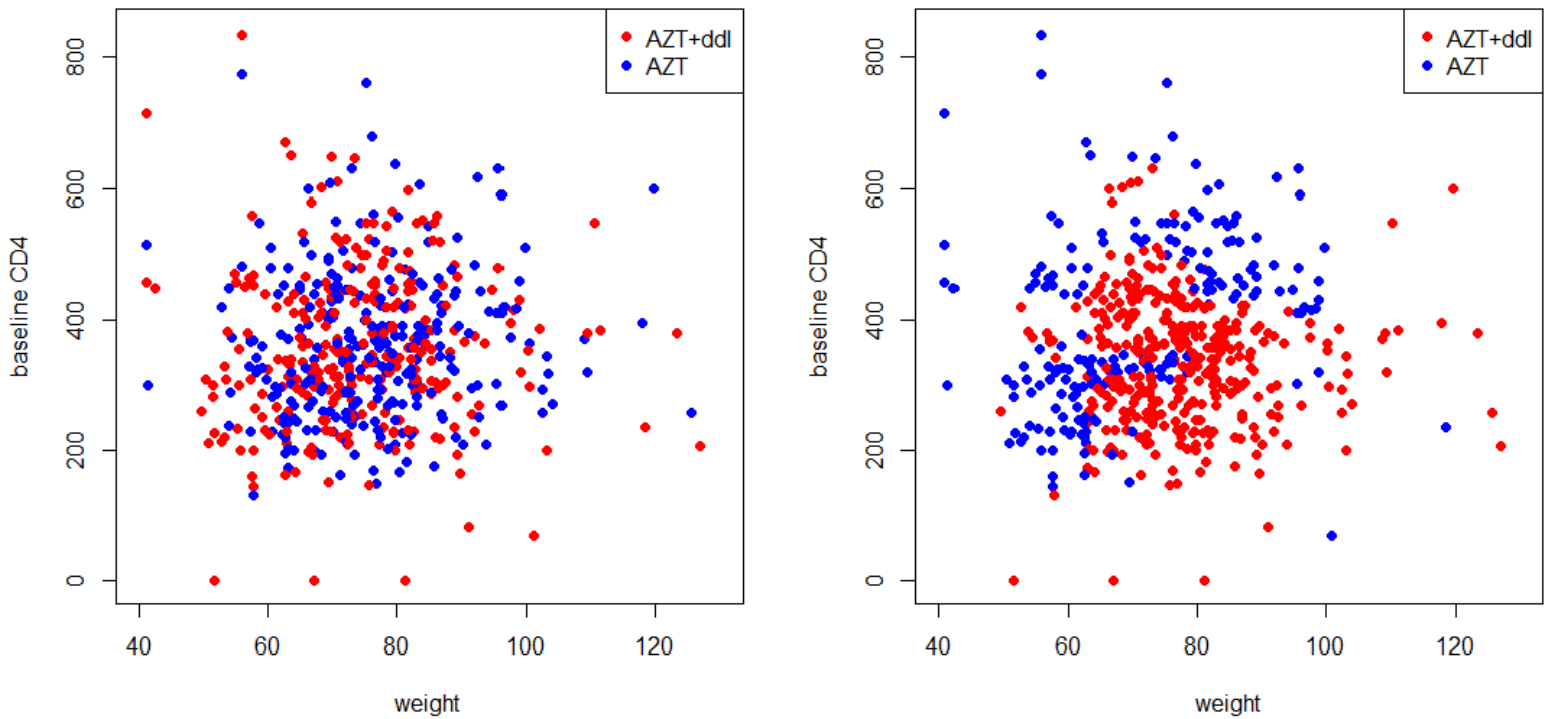

Figure 1: Original (left) and optimal (right) treatment assignment using QRF for 532 HIV-infected patients

4. Summary

Summary

Real-world data + machine learning + causal inference

- RWD and RWE provide complementary information for medical product development and regulatory decision making
- There are substantial challenges in deriving robust RWE from RWD due to confounding, either time-independent or time-dependent
- Real-world data (often big with three Vs), machine learning and causal inference methodologies play an increasingly important role in RWE development
  - Three "V" for big data: Volume, variety and velocity
  - Combining data from clinical trials and real-world studies
  - The seven-step journey of causal inference from RWD to RWE
- Back-door criteria to identify the set of covariates for confounding adjustment and front-door criteria to estimate causal effects
- From precision medicine to precision health to precision prevention

5. References

References (1/5)

1. Athey, S., G. W. Imbens, and S. Wager (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(4), 597–623.
2. Bagot, K., S. A. Matthews, M. Mason, L. M. Squeglia, J. Fowler, K. Gray, M. Herting, A. May, I. Colrain, J. Godino, et al. (2018). Current, future and potential use of mobile and wearable technologies and social media data in the abcd study to increase understanding of contributors to child health. *Developmental Cognitive Neuroscience 32*, 121–129.
3. Balzer, L., M. Petersen, and M. J. van der Laan (2016). Tutorial for causal inference. In P. Bhlmann, P. Drineas, M. Kane, and M. van der Laan (Eds.), *Handbook of Big Data*, pp. 361–386. Chapman & Hall/CRC Press.
4. Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.
5. Chen, J., J. Heyse, and T. L. Lai (2018). *Medical Product Safety Evaluation: Biological Models and Statistical Methods*. Chapman & Hall/CRC Press.
6. Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research 3*(Nov), 507–554.
7. Davenport, T. H. and R. Ronanki (2018). Artificial intelligence for the real world. *Harvard Business Review 96*(1), 108–116.
8. FDA (2018). Framework for FDAs real-world evidence program. US Food and Drug Administration, Silver Spring, MD (https://www.fda.gov/media/120060/download).
9. Fleming, N. (2018). Computer-calculated compounds. *Nature 557*(7707), S55–S57.
10. Ford, I. and J. Norrie (2016). Pragmatic trials. *New England Journal of Medicine 375*(5), 454–463.

INVENTING FOR LIFE

References (2/5)

11. Funk, M. J., D. Westreich, C. Wiesen, T. Stürmer, M. A. Brookhart, and M. Davidian (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology 173*(7), 761–767.
12. Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *Political Science*, 71–74. doi:10.1017/S1049096514001760.
13. Guo, R., L. Cheng, J. Li, P. R. Hahn, and H. Liu (2019). A survey of learning causality with data: Problems and methods. *arXiv preprint arXiv:1809.09337*.
14. Hammer, S. M., D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. Niu, et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine 335*(15), 1081–1090.
15. Hessler, G. and K.-H. Baringhaus (2018). Artificial intelligence in drug design. *Molecules 23*(10), 2520.
16. Kreif, N. and K. DiazOrdaz (2019). Machine learning in policy evaluation: new tools for causal inference. *arXiv preprint arXiv:1903.00402*.
17. Krittanawong, C., H. Zhang, Z. Wang, M. Aydar, and T. Kitai (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology 69*(21), 2657–2664.
18. Londhe, V. Y. and B. Bhasin (2018). Artificial intelligence and its potential in oncology. *Drug Discovery Today*.
19. Makady, A., A. de Boer, H. Hillege, O. Klungel, W. Goettsch, et al. (2017). What is real-world data? a review of definitions based on literature and stakeholder interviews. *Value in Health 20*(7), 858–865.

INVENTING FOR LIFE

References (3/5)

20. Monroe, B. L., J. Pan, M. E. Roberts, M. Sen, and B. Sinclair (2015). No! formal theory, causal inference, and big data are not contradictory trends in political science. *Political Science*, 71–74. doi:10.1017/S1049096514001760.

21. Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 331–355.

22. Nahum-Shani, I., M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy (2012). Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods 17*(4), 478–494.

23. Obermeyer, Z. and E. J. Emanuel (2016). Predicting the futurebig data, machine learning, and clinical medicine. *The New England Journal of Medicine 375*(13), 1216.

24. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press, New York. (Original edition, 2000).

25. Pearl, J., M. Glymour, and N. Jewell (2016). *Causal Inference in Statistics. A Primer*. John Wiley & Sons.

26. Pesapane, F., M. Codari, and F. Sardanelli (2018). Artificial intelligence in medical imaging: threat or opportunity? radiologists again at the forefront of innovation in medicine. *European Radiology Experimental 2*(1), 2–10.

27. Rajkomar, A., E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, et al. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine 1*(1), 18.

References (4/5)

28. Ramsey, J., M. Glymour, R. Sanchez-Romero, and C. Glymour (2017). A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International Journal of Data Science and Analytics 3*(2), 121–129.

29. Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, pp. 189–326. Springer.

30. Robins, J. M. and D. M. Finkelstein (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics 56*(3), 779–788.

31. Robins, J. M., M. A. Hernan, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*(5), 550–560.

32. Robins, J. M., M. A. Hernán, and U. Sirbert (2004). Effects of multiple interventions. In *Comparative Quantification of Health risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, Volume 1, pp. 2191–2230. Geneva: World Health Organization.

33. Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine 26*(1), 20–36.

34. Schmidt-Erfurth, U., A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunović (2018). Artificial intelligence in retina. *Progress in Retinal and Eye Research*.

35. Schulte, P. J., A. A. Tsiatis, E. B. Laber, and M. Davidian (2014). Q-and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science 29*(4), 640–661.

36. Schwarz, G. et al. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

References (5/5)

37. Shen, D., G. Wu, and H.-I. Suk (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering 19*, 221–248.
38. Shiffrina, R. M. (2016). Drawing causal inference from big data. *PNAS 113*(27), 7308–7309.
39. Shimizu, S., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research 7*(Oct), 2003–2030.
40. Shimizu, S., T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen (2011). Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research 12*(Apr), 1225–1248.
41. Silver, D., G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller (2014). Deterministic policy gradient algorithms. In *Proceedings of the 31st International Conference on Machine Learning*.
42. Spirtes, P., C. N. Glymour, R. Scheines, D. Heckerman, C. Meek, G. Cooper, and T. Richardson (2000). *Causation, prediction, and search*. MIT press.
43. Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction*. MIT press.
44. van der Laan, M. J. and A. R. Luedtke (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference 3*(1), 61–95.
45. van der Laan, M. J. and S. Rose (2018). *Targeted learning in data science: Causal inference for complex longitudinal studies*. Springer.
46. Vogt, M., S. Jasial, and J. Bajorath (2018). Extracting compound profiling matrices from screening data. *ACS Omega 3*(4), 4706–4712.
47. Wang, Z., T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas (2015). Dueling network architectures for deep reinforcement learning. *arXiv preprint arXiv:1511.06581*.
48. Zhang, B. and M. Zhang (2018). C-learning: A new classification framework to estimate optimal dynamic treatment regimes. *Biometrics 74*(3), 891–899.

INVENTING FOR LIFE