



HHS Public Access

Author manuscript

J Am Coll Radiol. Author manuscript; available in PMC 2019 March 01.

Published in final edited form as:

J Am Coll Radiol. 2018 March ; 15(3 Pt B): 563–568. doi:10.1016/j.jacr.2017.12.023.

Performing an Informatics Consult: Methods and Challenges

Alejandro Schuler, MS, Alison Callahan, PhD, Kenneth Jung, PhD, and Nigam H. Shah, PhD
Center for Biomedical Informatics Research, Stanford University, Stanford, California

Abstract

Our health care system is plagued by missed opportunities, waste, and harm. Data generated in the course of care are often underutilized, scientific insight goes untranslated, and evidence is overlooked. To address these problems, we envisioned a system where aggregate patient data can be used at the bedside to provide practice-based evidence. To create that system, we directly connect practicing physicians to clinical researchers and data scientists through an *informatics consult*. Our team processes and classifies questions posed by clinicians, identifies the appropriate patient data to use, runs the appropriate analyses, and returns an answer, ideally in a 48-hour time window. Here, we discuss the methods that are used for data extraction, processing, and analysis in our consult. We continue to refine our informatics consult service, moving closer to a learning health care system.

Keywords

Learning health system; observational study; practice-based evidence; clinical informatics

INTRODUCTION

Most medical decisions are made without the support of rigorous evidence [1,2] in large part due to the cost and complexity of performing randomized trials [3,4]. Even when guidelines exist, clinicians often do not have the time to read and understand them [1,5]. Furthermore, guidelines often do not apply to complex patients commonly seen in the clinic [1]. In practice, clinicians must use their judgment to make decisions, informed by their own experiences and the collective experience of their colleagues. This often leads to suboptimal care and creates waste and harm [6–8]. Increasingly, it is recognized that the clinical trial enterprise fails to produce relevant evidence for good clinical care [9].

Retrospective observational studies using the electronic health record (EHR) can generate evidence relevant to real patient populations [2]. We have operationalized that opportunity as an *informatics consult* that clinicians solicit the same way they would solicit other specialist consults. Obtaining a consult is a familiar process to clinicians and eliminates the friction between researchers and practitioners, ensuring that practice-based evidence is always readily available. Instead of sending one-way “reports,” we offer the consult as a dialogue

Corresponding author and reprints: Alejandro Schuler, MS, Center for Biomedical Informatics Research, Stanford University, 1265 Welch Road, Stanford, CA 94305; aschuler@stanford.edu.

The authors have no conflicts of interest related to the material discussed in this article.

between the clinician and consult team and among team data scientists so that we are not fooled by oddities in the data and obvious biases. The ultimate goal is to make use of all the evidence on hand to make the best possible decision for patient care.

For example, one clinician requested a consult to assess if the risk of diabetic eye disease is different in diabetic patients treated with rosiglitazone compared with diabetic patients not treated with rosiglitazone. In this case, completing the consult involved an iterative refinement of the analysis to determine an appropriate index time; in the end, we used onset of diabetes as the index time, and after basic matching on age, gender, and length of record, we determined that patients treated with rosiglitazone do not have a statistically significant difference in rate of diabetic eye disease compared with patients not treated with rosiglitazone.

As another example, we received a request from a hospitalist interested in the use of imaging after spinal fusion surgery. The hospitalist requested a consult to determine how many patients who underwent spinal fusion surgery also had a spinal x-ray performed during the inpatient stay when the surgery was performed and in the 2 weeks after surgery. We found that the majority of spinal fusion surgery patients had an x-ray taken during their inpatient stay, and fewer than 5% also had a second x-ray taken in the 2 weeks postsurgery.

The generation of good-quality evidence from observational data is not a trivial process, especially when operating on timescales corresponding to the course of care that unfold in days instead of publication schedules that span several months. All observational data are biased in terms of what population is observed (selection bias), what data are recorded on what patients (missing data), and what patients get what treatments (confounding). Depending on the question asked, different methods are required to extract the data, transform it into a useful form, and analyze it to produce evidence [10,11]. In many cases, the methods themselves are being actively researched, and questions remain about their implementation. Naturally, the operational details of the service, which are beyond the scope of this discussion, are equally important as the analysis methods used to generate evidence. We believe that despite these limitations, it is possible to offer a service that uses available data to produce the most up-to-date evidence possible and contextualize the findings for clinicians to incorporate in their decision making.

DATA EXTRACTION AND TRANSFORMATION

Data Sources and Infrastructure

Before beginning the analysis, an appropriate data set must be extracted from the EHR. In our consult, we use data from Stanford's EHR as well as from national claims data sets, such as Truven MarketScan, depending on the question at hand. Our data sets include both structured (eg, International Classification of Diseases, ninth rev codes) and unstructured (text) data. Text data are pre-processed with our text-processing workflow, which has been validated in multiple studies [12,13]. All data elements (e.g. procedures, diagnoses, note text, labs) are mapped to unique clinical concepts using a knowledge graph [14,15]. We anticipate soon having access to linked imaging data, which we will preprocess analogously to text data.

Before proceeding with the consult, we must determine whether we have data that are relevant to the question. We use the Stanford Advanced Temporal Language Aided Search (ATLAS) engine [16,17] to ensure that we have sufficient cohort sizes and data of the required modalities available to complete the consult. The ATLAS engine features a rich temporal query language that enables fast (subsecond response times) and powerful (simple commands define complex logical and temporal restrictions) searches over millions of patient records.

Phenotyping

To perform a search using ATLAS, we must determine the criteria that define the patients of interest (*phenotyping*) [18]. Improper phenotyping can create significant selection biases in the results of downstream analyses [19]. Phenotyping inherently requires domain knowledge because certain criteria may not be clearly or uniquely articulated in EHR data [11]. For instance, to find type 2 diabetic patients, one might search for any patients with an International Classification of Diseases, ninth rev diagnosis code of 250.00, or for patients with 3+ mentions of “t2dm” in their notes, or for patients who are on metformin and have a single mention of “diabetes.” Such “rules” to find diabetic patients are often referred to as *phenotyping algorithms*, and it is difficult to judge which is best without expert review [20,21]. We currently rely on the inquiring clinician to help us define an appropriate phenotyping rule for his or her consult.

Supervised machine learning is increasingly used for phenotyping. Instead of defining a rule, a small number of hand-labeled patients are used to train a model, which then classifies the remaining patients [22]. High-specificity rules may also be used to label the training patients [23]. These approaches lessen domain knowledge requirements and may reduce variability in the resulting cohorts. The volume of proxy signals in text and image data make these approaches attractive for labeling phenotypes that are not recorded as structured data in the EHR (eg, socio-economic variables) [14,24]. It may also be possible to include patients in the analysis cohort according to the model’s confidence in the phenotype assignment. We are investigating the use of these methods for our consult, but do not currently apply them.

Finally, because phenotype definitions are difficult to evaluate without expert-labeled data, stability analyses are a good way to detect potential biases. If there are multiple alternative phenotype rules or models, the same analysis should be performed using each of them and the final results compared.

Feature Construction

For analysis, each patient must be represented as a vector of *features* that describe their relevant clinical characteristics. The EHR contains many data modalities through time, including clinical notes, diagnoses, procedures, laboratory results, vitals, demographics, administrative data, and prescriptions. With learning algorithms that can handle huge numbers of variables, there is little reason not to exploit the richest possible set of features [25]. However, many choices can be made [26] (eg, representing diagnoses as counts or binary indicators, text as ontology-standardized term mentions or n-gram counts, or laboratory values or most extreme recent value as an indicator). Our knowledge graph also

allows us to aggregate features using domain knowledge to increase power, and multiple methods exist to do this [27,28]. In addition, each feature can be calculated at different points along the patient timeline, or the data could be filtered to include only recent measurements, among other possibilities [29,30]. It is also possible to engineer or learn composite features [31,32].

In our consult, we strike a balance using count representations of data from each modality without aggregation. As we integrate imaging data, we will consult with radiologists to decide on a standard set of image features to include. We continue to investigate whether richer representations lead to better analyses, and in what cases. As with phenotyping, stability analyses across multiple feature extraction schemes may reveal biases in causal analyses.

CLINICAL QUESTIONS AND THEIR ANALYSES

Before any analysis is begun, it is important to clearly define the kind of evidence that is sought [10]. Our consult service is staffed to support multiple types of analyses.

Descriptive and Exploratory Analyses

For complex patient populations, it is helpful to describe the population to see if interpretable patterns emerge. Traditionally, this has been done by computing summary statistics of some clinically important features across predefined subgroups of the population. Our consult can easily answer questions, such as “For individuals with cancer, what is the incidence rate of thyroiditis and adrenalitis?”

Unsupervised learning methods such as principle component analysis and hierarchical clustering may also be used for descriptive analyses, but it is unclear how to report their results to clinicians. These methods can help answer questions, such as “What are the different subtypes of pediatric autistic patients in our practice?” [33]. However, it is not clear how the results of an unsupervised analysis should be reported back in a way that appropriately trades off brevity for meaningfulness. The output of these methods is a representation of each patient in a low-dimensional space, but the dimensions of that space may not have an obvious clinical interpretation limiting their utility [34].

Inferential and Causal Analyses

Clinicians are often interested in the relationship between a treatment and an outcome. In an inferential analysis, it suffices to say that the outcome and treatment change together in a statistically significant way, but that does not imply that changing the treatment would change the outcome. To perform a *causal* analysis, we must attempt to reduce confounding while performing inference.

There are several challenges in using EHR data. First, EHR data are retrospective and observational, so highly prone to confounding; second, there are a variety of causal inference methods to choose from; third, it is difficult to evaluate if confounding has been addressed well enough; and fourth, causal effects vary between patients.

We attempt to control confounding using various strategies and are investigating methods to assess and report residual confounding. Causal inference methods exist to alleviate confounding biases [35]. Most of these methods work by matching subjects at baseline and using only the matched subsample for further analysis. Supervised learning can be used in some aspects of causal analyses (eg, in estimating propensity scores or predicting potential outcomes in matched cohorts) [36–38].

Typically, causal inference methods are chosen based on ease of use, researcher preference, or performance on simulated data [39,40]. Researchers experiment with various methods until they are satisfied with the result. Our consult must have a reasonably standardized methodological pipeline to return results quickly. Therefore, we run several analyses using established best practices with the most popular methods and report the results from each of them. We currently use an unadjusted regression, an adjusted regression after matching on common confounders [41], and an adjusted regression after 1:1 matching on a propensity score estimated from a large number of features. These three methods span the range between doing very little and doing a lot to address confounders and common best practice [35,42,43]. Because it is generally not possible to know which causal inference method will best estimate the true causal effect, we are currently developing a framework that takes a real data set and generates look-alike simulated data with a user-specified treatment effect. Preliminary work shows that methods that successfully estimate the user-specified effect in these simulated data sets also successfully estimate the true effect in the real data [44].

Simulations of this kind might help pick the right causal inference method but do not help assess the presence of unobserved confounders (or assess if confounding has been addressed well enough). *Negative control outcomes* are a useful mechanism to assess susceptibility to such biases [45]. Briefly, the studies are rerun using an alternative outcome that is known to not be caused by the treatment or comparator. If the results show a significant effect, then it is likely that the results of the original study are also biased. For example, a prescription of metformin could never be causative of preprescription HbA1c levels. Therefore, if we are interested in metformin's effect on postprescription HbA1c, we could use the preprescription value as a negative control. This procedure can be repeated with numerous negative controls to build a null distribution of effect sizes [19]. The challenge is finding appropriate negative controls [45]. We are considering crowdsourcing negative control outcomes from inquiring clinicians (ie, when they submit their question to the consult, they will also be asked to submit a reasonable negative control outcome relative to their true outcome and treatment). Over time, we will compile a database of negative control outcomes that will allow us to run multiple negative controls per study.

Often, the need to use our consult arises because of a particular patient. Currently, the inquiring physician outlines the cohort of “similar enough” patients for study. However, emerging methods use machine learning in conjunction with causal inference techniques to estimate a personalized treatment effect for each patient [46–48]. These techniques may remove the need of quantifying patient similarity for cohort construction. We are exploring these methods for future use.

Predictive Analyses

Unlike inferential and causal analyses, *predictive* analyses do not seek structural knowledge about the world. Instead, they exploit correlations to predict outcomes or classify patient states at the individual level. Methodologically, predictive analyses are a straightforward application of supervised learning. Their results are evaluated based on their ability to correctly predict or classify previously unseen patients. Without further inference, model internals (eg, coefficients) do not provide mechanistic knowledge and should not be interpreted [49].

Clinically, predictive models can be used to make prognoses for new patients, as long as their treatment patterns do not deviate statistically from those observed in the training data. If prognostic models are used to inform decision making that affects outcomes, their predictions are no longer valid because new patients may be treated differently from those in the training data [50]. This creates a tension in the use of prognostic models. Without making assumptions or using causal inference techniques, prognostic models should not be used for decision making that affects the treatment choice [51]. When clinicians ask, “What will the outcome of this patient be?,” they are often actually interested in the causal question “What will the outcome be on this treatment versus on another?”

Predictive models are also useful for identifying latent disease. In this case, the outcome or condition has already occurred, so the use of predictive models is always statistically sound. Diagnostic models can be used to answer questions such as “Does my patient have familial hyperlipidemia?” However, care must be taken to define the time at which a patient is considered to have the condition. A model trained to discern long-term diabetics from healthy patients will perform poorly in diagnosing new diabetic patients and many of the important “predictors” of the disease will be treatments indicated by the disease [52].

Despite these caveats, we aim to provide predictive analyses through our consult. Method selection for predictive analyses is more straightforward than for causal analyses. We plan to employ a variety of machine learning methods (eg, regularized regression, tree ensembles, and neural nets) and use cross-validation to choose among them. We would create a predictive model for the cohort and report the model’s accuracy and calibration (or mean squared error) along with the prediction on the patient in question. Questions remain about which accuracy metrics to report to clinicians, especially in diagnostic cases where, for example, a false-negative would be more costly than a false-positive.

DISCUSSION AND CONCLUSION

Our informatics consult is capable of answering different kinds of clinical questions with state-of-the-art analysis methods, but many methodological issues in data extraction, transformation, and analysis are still being researched.

The first time such an idea to use aggregated patient data for decision making was put in play was in 1972 [53]. Much progress has been made in analysis methods, search technology, and data availability that make our current service possible. However, there are other issues unrelated to the choice of analysis methods that affect the final feasibility and

utility of offering such a service. These include the funding model for such a consult service, assessment of the risk-benefit of using an on-demand evidence generation service, and the possibility that the “answer” changes as more data come in.

Our informatics consult service attempts to create a synergy between a thorough study using sound methods and clinical judgment, enabling the rapid generation of applicable clinical evidence where there was none before. The informatics consult is a step towards a fully integrated learning health care system.

References

1. Stewart WF, Shah NR, Selna MJ, Paulus RA, Walker JM. Bridging the inferential gap: the electronic health record and clinical evidence. *Health Aff.* 2007; 26:w181–91.
2. Longhurst CA, Harrington RA, Shah NH. A “green button” for using aggregate patient data at the point of care. *Health Aff.* 2014; 33:1229–35.
3. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ.* 1996; 312:1215–8. [PubMed: 8634569]
4. Eapen ZJ, Vavalle JP, Granger CB, Harrington RA, Peterson ED, Califf RM. Rescuing clinical trials in the United States and beyond: a call for action. *Am Heart J.* 2013; 165:837–47. [PubMed: 23708153]
5. Druss BG, Marcus SC. Growth and decentralization of the medical literature: implications for evidence-based medicine. *J Med Libr Assoc.* 2005; 93:499–501. [PubMed: 16239948]
6. Institute of Medicine, Roundtable on Evidence-Based Medicine. The learning healthcare system: workshop summary. Washington DC: National Academies Press; 2007.
7. Smith, M.Saunders, R.Stuckhardt, L., McGinnis, JM., editors. Committee on the Learning Health Care System in America, Institute of Medicine. Best care at lower cost: the path to continuously learning health care in America. Washington DC: National Academies Press; 2014.
8. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med.* 2014; 174:710–8. [PubMed: 24663331]
9. Caruso, C. Robert Califf: “The clinical trials enterprise has gone awry”. STAT. Available at: <https://www.statnews.com/2017/06/21/robert-califfs-clinical-trials/>. Published June 21, 2017. Accessed June 26, 2017
10. Leek JT, Peng RD. What is the question? *Science.* 2015; 347:1314–5. [PubMed: 25721505]
11. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* 2013; 51:S30–7. [PubMed: 23774517]
12. LePendu P, Iyer SV, Fairon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. *J Biomed Semantics.* 2012; 3(Suppl 1):S5.
13. Jung K, LePendu P, Iyer S, Bauer-Mehren A, Percha B, Shah NH. Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. *J Am Med Inform Assoc.* 2015; 22:121–31. [PubMed: 25336595]
14. LePendu P, Iyer SV, Bauer-Mehren A, et al. Pharmacovigilance using clinical notes. *Clin Pharmacol Ther.* 2013; 93:547–55. [PubMed: 23571773]
15. Noy NF, Shah NH, Whetzel PL, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.* 2009; 37:W170–3. [PubMed: 19483092]
16. Banda, JM., Callahan, A., Kale, D., Polony, V., Shah, NH. Advanced temporal language aided search for the OHDSI community. Available at: http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=resources:ohdsi-submission-template_2016_atlas1_ac.pdf. Accessed January 6, 2018
17. Shah, Nigam. Search engine. Youtube. 2017. Available at: <https://www.youtube.com/watch?v=HUUp04tA8BM>. Accessed January 6, 2018
18. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc.* 2013; 20:e147–54. [PubMed: 23531748]

19. Madigan D, Stang PE, Berlin JA, et al. A systematic statistical approach to evaluating evidence from observational studies. *Annu Rev Stat Appl.* 2014; 1:11–39.
20. Peissig PL, Rasmussen LV, Berg RL, et al. Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *J Am Med Inform Assoc.* 2012; 19:225–34. [PubMed: 22319176]
21. Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. *J Am Med Inform Assoc.* 2013; 20:e319–26. [PubMed: 24026307]
22. Yu S, Liao KP, Shaw SY, et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. *J Am Med Inform Assoc.* 2015; 22:993–1000. [PubMed: 25929596]
23. Agarwal V, Podchyska T, Banda JM, et al. Learning statistical models of phenotypes using noisy labeled training data. *J Am Med Inform Assoc.* 2016; 23:1166–73. [PubMed: 27174893]
24. Poplin, R., Varadarajan, AV., Blumer, K., et al. Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. arXiv:1708.09843 [cs.CV]. Available at: <https://arxiv.org/abs/1708.09843>. Accessed January 6, 2018
25. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2017; 24:198–208. [PubMed: 27189013]
26. Ozery-Flato M, Yanover C, Gottlieb A, Weissbrod O, Parush Shear-Yashuv N, Goldschmidt Y. Fast and efficient feature engineering for multi-cohort analysis of EHR data. *Stud Health Technol Inform.* 2017; 235:181–5. [PubMed: 28423779]
27. Winnenburt R, Shah NH. Generalized enrichment analysis improves the detection of adverse drug events from the biomedical literature. *BMC Bioinformatics.* 2016; 17:250. [PubMed: 27333889]
28. Pivovarov R, Elhadad N. Automated methods for the summarization of electronic health records. *J Am Med Inform Assoc.* 2015; 22:938–47. [PubMed: 25882031]
29. Zhao J, Papapetrou P, Asker L, Boström H. Learning from heterogeneous temporal data in electronic health records. *J Biomed Inform.* 2017; 65:105–19. [PubMed: 27919732]
30. Tran T, Luo W, Phung D, et al. A framework for feature extraction from hospital medical data with applications in risk prediction. *BMC Bioinformatics.* 2014; 15:425. [PubMed: 25547173]
31. Choi, E., Bahadori, MT., Searles, E., Coffey, C., Sun, J. Multi-layer representation learning for medical concepts. arXiv [cs.LG]. Available at: <http://arxiv.org/abs/1602.05568>. Accessed January 6, 2018
32. Wang G, Jung K, Winnenburt R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. *J Am Med Inform Assoc.* 2015; 22:1196–204. [PubMed: 26232442]
33. Stevens MC, Fein DA, Dunn M, et al. Subgroups of children with autism by cluster analysis: a longitudinal examination. *J Am Acad Child Adolesc Psychiatry.* 2000; 39:346–52. [PubMed: 10714055]
34. Schuler A, Liu V, Wan J, et al. Discovering patient phenotypes using generalized low rank models. *Pac Symp Biocomput.* 2016; 21:144–55. [PubMed: 26776181]
35. Stuart EA, DuGoff E, Abrams M, Salkever D, Steinwachs D. Estimating causal effects in observational studies using Electronic Health Data: challenges and (some) solutions. *EGEMS (Wash DC).* 2013; 1 <https://doi.org/10.13063/2327-9214.1038>.
36. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin Epidemiol.* 2010; 63:826–33. [PubMed: 20630332]
37. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol.* 2017; 185:65–73. [PubMed: 27941068]
38. Hill JL. Bayesian nonparametric modeling for causal inference. *J Comput Graph Stat.* 2011; 20:217–40.
39. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol Drug Saf.* 2008; 17:546–55. [PubMed: 18311848]
40. Colson KE, Rudolph KE, Zimmerman SC, et al. Optimizing matching and analysis combinations for estimating causal effects. *Sci Rep.* 2016; 6:23222. [PubMed: 26980444]

41. Castro VM, Apperson WK, Gainer VS, et al. Evaluation of matched control algorithms in EHR-based phenotyping studies: a case study of inflammatory bowel disease comorbidities. *J Biomed Inform.* 2014; 52:105–11. [PubMed: 25196084]
42. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat.* 2011; 10:150–61. [PubMed: 20925139]
43. Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Stat Med.* 2014; 33:1057–69. [PubMed: 24123228]
44. Schuler A, Jung K, Tibshirani S, Hastie T, Shah N. Synth-validation: Selecting the best causal inference method for a given dataset 2017. arXiv preprint arXiv:1711.00083.
45. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology.* 2010; 21:383. [PubMed: 20335814]
46. Athey, S., Imbens, G. Recursive partitioning for heterogeneous causal effects. arXiv [stat.ML]. 2015. Available at: <https://arxiv.org/abs/1504.01132>
47. Wager, S. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association.* <https://arxiv.org/abs/1510.04342> [stat.ME]
48. Powers, S., Qian, J., Jung, K., Schuler, A., Shah, NH., Hastie, T., et al. Some methods for heterogeneous treatment effect estimation in high-dimensions. *Statistics in Medicine.* <https://arxiv.org/abs/1707.00102> [stat.ML]
49. Shmueli G. To explain or to predict? *Stat Sci.* 2010; 25:289–310.
50. Hartford, J., Lewis, G., Leyton-Brown, K., Taddy, M. Counterfactual prediction with deep instrumental variables networks. arXiv:1612.09596 [stat.AP]. 2016. Available at: <https://arxiv.org/abs/1612.09596>. Accessed January 10, 2018
51. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. Intelligent models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10–13 2015; Sydney Australia: ACM; 2015.* p. 1721-1730.
52. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. *JAMA.* 2016; 316:1818–9. [PubMed: 27802529]
53. Feinstein AR, Rubinstein JF, Ramshaw WA. Estimating prognosis with the aid of a conversational-mode computer program. *Ann Intern Med.* 1972; 76:911–21. [PubMed: 4554880]

TAKE-HOME POINTS

- Our informatics consult connects clinicians with researchers capable of answering different kinds of clinical questions with state-of-the-art analysis methods, but many methodological issues in data extraction, transformation, and analysis are still being researched.
- We use data from multiple modalities (eg, codes, laboratory results, texts) to create more accurate patient representations.
- We use a search engine to quickly build rule-based cohorts and explore our data.
- We strike a balance between efficiency and rigor when performing inferential analyses.
- Exploratory and predictive analyses are difficult to standardize and may be of limited actionability.
- We are working to implement and compare novel inference methods to the standard methods we currently use.