MICHIGAN ROSS

Big Data and the Precision Medicine Revolution

Wallace J. Hopp
Stephen M. Ross School of Business
University of Michigan

Jun Li
Stephen M. Ross School of Business
University of Michigan

Guihua Wang
Stephen M. Ross School of Business
University of Michigan

UNIVERSITY OF MICHIGAN

Big Data and the Precision Medicine Revolution

Wallace J. Hopp
Ross School of Business
University of Michigan
Ann Arbor, MI 48109
whopp@umich.edu
734-763-6027
Corresponding author

Jun Li
Ross School of Business
University of Michigan
Ann Arbor, MI 48109
junwli@umich.edu
734-763-4612

Guihua Wang
Ross School of Business
University of Michigan
Ann Arbor, MI 48109
guihuiw@umich.edu

Abstract

The big data revolution is making vast amounts of information available in all sectors of the economy including health care. One important type of data that is particularly relevant to medicine is observational data from actual practice. In comparison to experimental data from clinical studies, observational data offers much larger sample sizes and much broader coverage of patient variables. Properly combining observational data with experimental data can facilitate precision medicine by enabling detection of heterogeneity in patient responses to treatments and tailoring of health care to the specific needs of individuals. However, because it is high-dimensional and uncontrolled, observational data presents unique methodological challenges. The modeling and analysis tools of the production and operations management field are well-suited to these challenges and hence POM scholars are critical to the realization of precision medicine with its many benefits to society.

Keywords: big data, precision medicine, observational data, machine learning, causal inference

1. Introduction

Two parallel revolutions are about to transform the health care industry in profound ways. The first is in the practice of medicine, where imprecise one-size-fits-all medicine is being replaced by individually tailored precision medicine. The second is the analysis of data, where big data techniques are making it possible to extract patterns from vast amounts of digital information to guide decision making in a wide range of sectors and applications. Together these revolutions will fundamentally alter the clinical practice of medicine, by changing the way patients are diagnosed, treatments are selected, and care is delivered. But they will also have ramifications for the health care industry far beyond the practice of medicine, including the way patients choose and interact with providers, the way providers make strategic and tactical decisions, and the way payers reimburse and incentivize both patients and providers.

The link between these two revolutions is data. The big data revolution is producing new ways to generate and analyze data and the precision medicine revolution is creating new ways to leverage it (see Kruse et al. (2016) for a broad review of the medical literature on using data in health care). In particular, big data methods are making it possible to go beyond experimental data generated by clinical studies and take advantage of observational data about patients in uncontrolled settings. Observational data can come from medical records, online reviews, mobile devices and many other sources. In contrast to experimental data, observational data is cheaper to obtain, has much larger sample sizes and addresses a much wider range of variables. For example, less than 5% of adult cancer patients are part of a clinical trial (Unger et al., 2017). Furthermore, since only patients meeting certain criteria can enroll in trials, patients with uncommon tumor types, older patients, and patients with a poor performance status or comorbidities are frequently underrepresented. As a result, researchers are increasingly looking to observational data to study cancer treatments and to identify side effects not apparent in randomized clinical trials (Spigel, 2010). By making virtually every act of health maintenance and patient treatment available for study, big data approaches are opening the possibility of true precision medicine that enables patients to receive care best matched to their specific health condition, individual characteristics and personal preferences.

However, although observational data has enormous potential, it presents problems not present with experimental data. In clinical experiments subjects are selected to minimize bias and are organized into treatment and control groups, which makes for crisp statistical conclusions. In contrast, observational data comes from actual patient experiences and can therefore suffer from bias, censoring, insufficient sample size and other problems. Therefore, in order to use observational data to guide health care decisions we must find ways to correct for the inherent flaws in the data. This is where the emerging tools of analytics and machine learning come in, and where the POM field can play an important role in bringing about the merger of big data and precision medicine to revolutionize health care.

In the following, we describe the path to precision medicine, the roles new types of data will play, the methodological challenges of collecting and using these data, and the stakeholder impacts and benefits that will ensue. We pay particular attention to research streams and opportunities where POM scholars are having, and can continue to have, a major impact on the big data/precision medicine revolution.
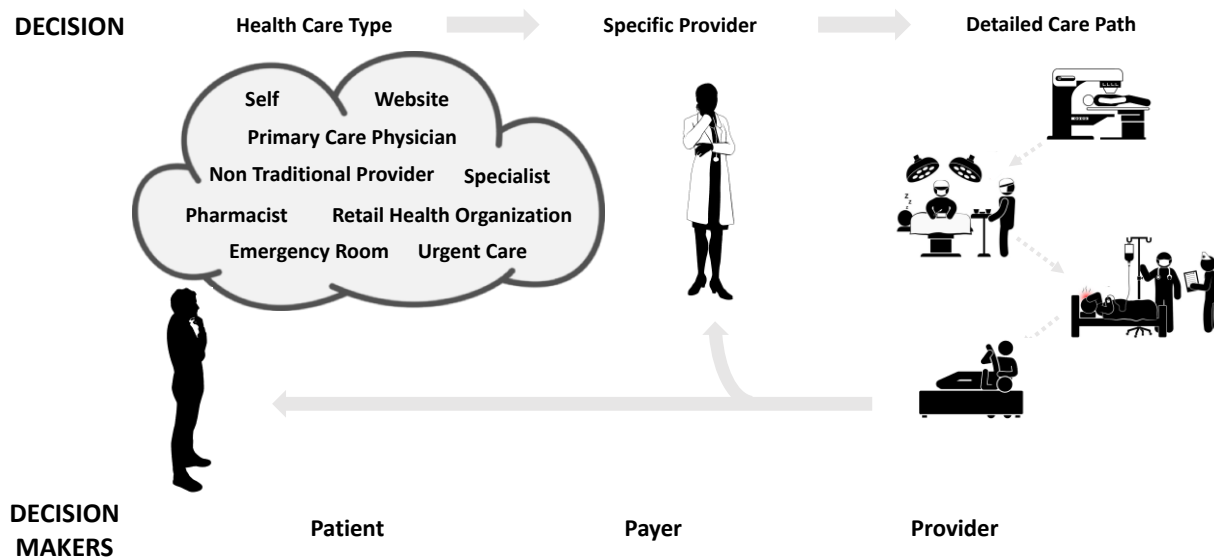
2. Precision Medicine

In stark contrast with one-size-fits-all medicine, in which disease treatment and prevention strategies are applied to everyone without regard to individual differences, precision medicine involves "the tailoring of medical treatment to the individual characteristics of each patient" (National Research Council, 2011). The Federal Drug Administration (2018) elaborated on this distinction with the following description: "Most medical treatments are designed for the "average patient" as a "one-size-fits-all-approach," which may be successful for some patients but not for others. Precision medicine, sometimes known as "personalized medicine", is an innovative approach to tailoring disease prevention and treatment by

2

taking into account differences in people's characteristics, environments, and lifestyles. The goal of precision medicine is to target the right treatments to the right patients at the right time."

Precision medicine, like all health care, involves two basic elements – decision making and execution. We must decide what to do, and then do it properly. Of course, execution of the various health care activities (tests, procedures, medications, etc.) is essential. But precision medicine is concerned primarily with selecting the right activities for a given patient, i.e., the decision-making facet of health care. Most discussions of precision medicine focus specifically on decisions made by a clinician regarding treatment (e.g., what drug/dosage to prescribe).

But health care decisions involve much more than medical treatment and clinicians. As we depict in Figure 1, the health care process is usually initiated by a patient selecting a type of care. This could range from consulting a website for self-treatment information to deciding to see a medical specialist. Then the patient, possibly with assistance from a health care professional or influence from an insurance company, selects a provider. Again, there is a wide range of options, from engaging a fitness coach to selecting a surgeon. Finally, the provider, hopefully with strong input from the patient and probably with awareness of payer policies, implements some kind of care path. This could involve tests, medications, therapies or any number of other measures to address the patient's needs. Frequently, because current health care is far from precision medicine, iteration is required. The patient may try multiple types of health care and engage multiple providers. The locus of decision making may shift between the patient, provider and payer, with the patient dominating type of health care decisions, providers dominating care path decisions and the payer influencing decisions across the spectrum. Whether or not the patient winds up satisfied with his/her health outcome depends on the combined impact of all these decisions, as well as the quality of execution within the care paths.

Figure 1: Health Care Decisions.

To achieve its full potential, precision medicine must address all of these levels of heath care decisions. It should help patients make good decisions about what type of care to seek. For example, a diabetic patient should be able to get information about what symptoms indicate he/she should seek diet advice on Dlife.com and what symptoms indicate a need for prompt medical intervention. Precision medicine should also help patients (and providers) choose the right provider for a particular patient. For example, a

patient in need of heart surgery, should have access to rankings or outcome data to help him/her (perhaps with help from his/her cardiologist) choose a surgeon. Finally, precision medicine should help all of the actors involved in implementing a care path make the best decisions about tests, procedures, dosages, and other aspects of the health care process.

Although "precision medicine" has emerged as the most popular term for individually tailored health care, there are other terms in use. The older "personalized medicine" is regarded by many as interchangeable with personalized medicine, but by some as a literal call for individualized medicine (see National Research Council, 2011). Since, in some cases, the most effective treatment may not be individually tailored, these people argue (without excuse for the pun) that "precision medicine" is the more precise term. "Pharmacogenomics" is a specific instance of precision medicine, which makes use of genetic information to tailor pharmacological treatments to patients. Finally, an overlapping term is "evidence-based medicine" which refers to the use of relevant data to guide medical decisions. However, while precision medicine is always evidence-based, not all evidence-based medicine is personalized. As we describe in Wang et al. (2017), many research studies focus on the response of an average patient to a particular treatment, and therefore imply "one-size-fits-all" protocols.

3. The Role of Big Data in Precision Medicine

Neither one-size-fits-all medicine, nor precision medicine, nor the stages in between them, are uniquely defined entities. Instead they represent a continuum of practices whose outcomes improve with the amount and type of data they leverage to guide decisions.

The continuum of one-size-fits-all medicine ranges from a pure trial-and-error process, in which the patient (or physician) selects randomly from a set of treatments in hopes of achieving an improvement, to evidence-based priority rules, in which the patient (physician) tries alternatives in decreasing order of their probability of success. By using success rate data to rank order the alternatives, we can increase the likelihood and speed of achieving a good outcome. More accurate success rate data will produce a more efficient search process. However, regardless of whether success rates are raw or risk adjusted[1], this type of data will lead to the same rank ordering of alternatives for all patients, and hence one-size-fits-all medicine. That is, all patients will start with the top-ranked option and work their way down the list in the same order. When patient responses to the various treatments are heterogeneous, this ensures inefficiency in the process, no matter how refined the success rate data.

Precision medicine can also involve trial-and-error, but the treatment alternatives and/or likelihoods of success are tailored to the individual patient. The most basic form of precision medicine would be a treatment that is determined by a single individual characteristic (e.g., a dosage based on the precise weight of the patient). As precision medicine takes into account more patient descriptors, such as comorbidities, genetic markers, preferences, lifestyle attributes, etc., the alternative set and outcome likelihoods can be refined. One reason pharmacogenomics has received so much attention within the precision medicine literature is that scientists expect many new indicators of patient responses to drugs to be discovered at the genetic level.
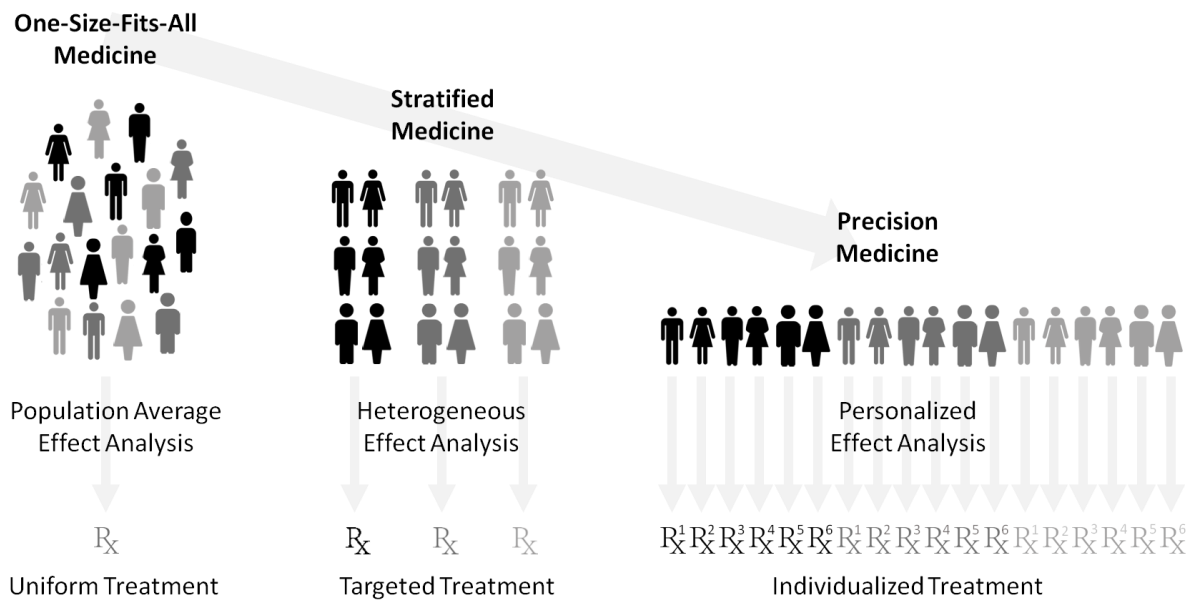
Between one-size-fits-all and precision medicine is stratified medicine, which divides patients into groups according to their response to a therapy. Since there may be other differences between patients in the

---

[1] Risk-adjusted success rates correct for differences in patient characteristics by computing the average outcome from each alternative for a mix of patients that reflects the population. Note that the expected outcome for a "population average" patient may not reflect the outcome for any actual patient. Hence, as we note in Wang et al. (2017), population average data may not give appropriate rankings of alternatives when patient outcomes are heterogeneous.

same group, treatments will be rank ordered according to in-group success rates. Hence, we can think of this as "one-size-fits-some" medicine. An example of stratification is the classification of breast cancers into estrogen receptor (ER) positive and negative categories. ER positive cancers respond to hormone therapies, while ER negative cancers do not. Therefore, women in the two categories receive different chemotherapy protocols. Stratified medicine is also a continuum of practices because additional characteristics can be used to refine patient groupings. For example, breast cancers can also be classified as progesterone receptor (PR) positive or negative and as human epidermal growth factor receptor 2 (HER2) positive or negative. Because patients in these categories respond differently to drugs, chemotherapy protocols can be better targeted by using the PR and HER2 biomarkers. When indicators get specific enough to divide patients into groups of one, stratified medicine becomes precision medicine.

Figure 2 depicts the continuum from one-size-fits-all to precision medicine. For our purposes of linking the precision medicine revolution to the big data revolution, the most important aspect of this figure is the nature of the data analysis required to facilitate each type of medicine.

Figure 2: Progression toward Precision Medicine.

One-size-fits-all medicine requires *population average effect analysis*. The simplest form of such analysis is the frequentist approach, which uses outcomes to calculate the fraction of the population for which each alternative is effective. But when health outcomes cannot be measured in binary success/failure terms (e.g., liver transplant patients are evaluated in graft survival time, a continuous metric), a more sophisticated approach is required. In these circumstances, a risk-adjusted outcome analysis can be used to rank alternatives. For example, to compare the effectiveness of different hospitals in treating a particular cardiovascular disease, we can evaluate each hospital by dividing the outcome (e.g., five-year survival rate) by the expected outcome (i.e., the average five-year survival rate that would be achieved by the full population of hospitals if they treated the same mix of patients as the individual hospital). The expected outcome requires a regression across all patients and hospitals to compute (see Glance, et al., 2006). The resulting observed-to-expected, or O/E, ratio provides a single number metric with which to rank providers. However, a single rank ordering based on O/E ratio is only appropriate when the relative

effectiveness of the providers does not depend on patient characteristics. When it does, rank order may be incorrect for some, or even all, patients.

Stratified medicine uses *heterogeneous effect analysis* to correct for the failure of average effect analysis to account for patient differences. This requires patient characteristics along with outcomes. The more characteristics we can observe (e.g., age, gender, comorbidities, biomarkers, readings from wearable devices, etc.), the more likely we are to find characteristics that differentiate patients with regard to their responses. However, the more finely we stratify patients, the smaller our sample sizes become. In clinical studies, where data collection is expensive and difficult, it is often impossible to achieve large enough samples to permit subgroup analyses needed to detect heterogeneous patient responses.[2] As we discuss below, finding the right balance between difference distinction and statistical significance is a key analytical challenge in implementing stratified medicine.

Precision medicine uses *personalized effect analysis*, which often requires personalized data. Genome sequencing data is one form of personalized data. But so are preference (e.g., risk aversion) and lifestyle (e.g., diet and sleep) data. One could argue that all medicine makes use of at least some personalized data in the form of patient responses to provider inquiries (e.g., "How do you feel about the proposed course of action?). But precision medicine is aimed at a much more objective, evidence-based incorporation of personalized data than these traditional subjective assessments. Some of these data are being collected through traditional channels (e.g., doctor appointments) and compiled in electronic medical records (EMRs) or electronic health records (EHRs). But more detailed personalized data will require new approaches, such as mining social media or using handheld devices.

A characteristic common to the data used for above analyses is multidimensionality. Because treatment alternatives are generally evaluated along multiple criteria, such as expected clinical outcome, risk of side effects, comfort and convenience of care, cost, and others, we need data with multiple dependent variables, as well as multiple independent variables, to support decision making. We can use big data to estimate the heterogeneity of patient responses along each dimension. But this will not be enough to facilitate patient treatment choices. We will also need some way to combine the results for the various dimensions. To do this at the personalized level, we will need individual preference data (in the form of linear weights or parameters for a nonlinear utility function). Furthermore, preferences on non-medical dimensions, such as travel distance, expense, familiarity, support services for family members, and many others, may be relevant to individual decisions. By combining heterogeneous outcome data and individual patient preference data we can achieve true precision medicine, in which choices are customized to the level of the individual patient.

Ultimately, the progression from one-size-fits-all to precision medicine will be driven by more and better sources of data, and the analytics techniques needed to use them to guide health care decisions. But collecting and analyzing these data present a wide array of challenges, as we describe below.

4. Challenges in Collecting and Analyzing Big Data

Precision medicine can make use of both experimental and observational data. But, although challenges remain in the collection and use of experimental data to support health care decisions, these are familiar

---

[2] For example, Park et al. (2015) compared percutaneous coronary intervention (PCI) using stents with coronary-artery bypass grafting (CABG) in treating multi-vessel coronary artery disease. Although they found the rate of major adverse cardiovascular events was significantly higher among PCI patients than CABG patients, their ability to detect differences in outcomes among patient subgroups was limited by the size of the study. The study involved a total of 880 patients, which permitted analysis of only a small number of patient characteristics and restricted the statistical power of the analyses that were done.

problems clinical researchers have been dealing with for years. The new challenges presented by the incorporation of big data into health care decision making, and the ones most amenable to the skills of the POM community, are those associated with observational data. In this section, we first describe collection issues associated with this increasingly large-scale and personalized type of data. We then discuss estimation issues related to high dimensionality, bias, censoring and sample size issues that are common in uncontrolled observational data. Next, we highlight some modeling and optimization challenges that must be addressed to translate observational data into decision aids. Finally, because the ultimate goal is to enable all stakeholders in the health care system to make better decisions, we describe implementation challenges involved in presenting and disseminating information derived from big data analyses.

4.1 Data Collection

Health care observational data comes from many sources and in many formats. In this subsection, we describe common challenges that arise in the process of collecting large-scale observational data. These include technological barriers, incomplete or inaccurate data, privacy concerns and scientific challenges.

4.1.1 Technological Barriers

Advances in Information Technology (IT) have transformed health care data collection – from hand written notes, electronic health records, personal wearable devices, and many other sources. These advances also made it possible to capture both structured (e.g., administrative and claims data) and unstructured data (e.g., texts, images, audios, videos). While structured data are straightforward to analyze, and some text can be analyzed through textual analysis (see e.g., Wu, 2016), technology for converting medical images, audios and videos to textual or structured data that can be easily understood by non-medical researchers is still lacking. As a result, for example, echocardiograms of mitral patients are routinely interpreted by a cardiologist or surgeon but are rarely are rarely available for use by a data analyst or policy maker.

There is a lack of coordination of health IT implementation across different organizations or even in different departments of the same organization. Variations in data collection platforms and storage methods make it difficult to convert data from different sources into the same format for analysis. The lack of coordination across different organizations is a major barrier to health information exchange when a patient switches to a different health care provider or payer. For example, when a patient changes insurance company, a new enrollee ID is usually created. Duplicate IDs make it hard to link data from multiple insurance companies to the same patient.

IT companies and researchers are continually seeking ways to translate complex and unstructured forms of information into usable data and hospitals and their IT providers are constantly working on better integration and interoperability. But both of these have a great deal of room for improvement. As a result, researchers seeking to use observational data to refine precision medicine must work with less-than-ideal data sets and are likely to do so for some time. But new research opportunities will continue to arise as these technological challenges are addressed.

4.1.2 Incomplete or Inaccurate Data

Missing values are common in observational data. These can occur due to unavailability of information (e.g., because a patient refused to disclose personal information or tests were not done), omissions by data administrators, accidental deletion, or data loss during merge and conversion. Missing values can be categorized broadly into three types: missing completely at random (i.e., independent of both the outcome variable and other independent variables), missing at random (i.e., independent of either the outcome variable or other independent variables), and missing not at random (i.e., correlated with the outcome

variable even after controlling for other independent variables). As we will discuss later, missing at random or completely at random may not be an important issue but missing not at random could cause biases in estimation. Techniques for dealing with missing value include imputation (i.e., replacing missing values with substituted values), deletion, interpolation and full analysis (see Little and Rubin (2014) for more details).

Worse than missing values are inaccurate values. Inaccurate values can be the result of failure to understand terminology, over-reliance on manual input, inconsistent coding practice, or change of documentation policies. Compared with missing values, inaccurate values are much more difficult to detect, especially when they are not obviously outliers. One approach to identifying inaccurate values is to compare different data (e.g., claims, clinical and administrative data) for the same information. Another is to verify the raw data used to calculate the values. For example, one might check a patient's height and weight if the patient's body mass index has a suspicious value.

4.1.3 Privacy Concerns

Privacy is becoming an increasingly serious concern for patients as individual sensors have made it possible to monitor health at an increasingly detailed level. Users may be unwilling to share their personalized data with the providers of these devices or their clients. This could hinder prospects for collecting vast amounts of personalized data to be used in precision medicine as well as developing new and more accurate devices for health monitoring. Encryption and de-identification are traditional approaches to addressing privacy concerns. Medicalchain is a recently developed blockchain approach to protecting patient information such as those from EMRs while sharing data across medical providers and treatment sites (Azaria et al., 2016).

Privacy is a concern to health care providers and payers, as well as to patients. Providers may be unwilling to share clinical data because there are laws and regulations that protect the privacy of patient information. They are also concerned that sharing a best practice with other providers may erode their competitive advantage in attracting patients. Payers may be unwilling to share their claims data because the data usually contain sensitive information about their cost structure and financial performance. Finally, providers and payers may be unwilling to share information with each other, because they do not want to lose negotiating power. In theory, encryption and de-identification can be used to protect provider and payer data. But doing this while enabling integration of data across participants and platforms is difficult. Data science scholars (see, e.g., Li and Qin, 2017, Miller and Tucker, 2017) have proposed to use encryption and other means for protecting patient privacy when medical records are shared.

4.1.4 Scientific Challenges

Pharmacogenomics makes use of genetic information to predict patient responses to medication. Recent advances in genetic technology offer the prospect of being able to personalize medications to patients based on genetic tests. However, while an increasing number of biomarkers have been discovered, their influence on outcomes is complex. As a result, few biomarkers are currently used in clinical decision making or drug development and their ultimate utility is still a matter of debate.

Several scientific challenges make it difficult to collect useful and accurate pharmacogenomics data. First, there are potentially many yet-to-be-identified genes and biomarkers that affect the pharmacology of a drug. Second, even for the identified pharmacogenomics biomarkers, some of the clinical results regarding the genetic association between the biomarker and the pharmacology of a drug have been inconsistent (Lam, 2013). Third, in addition to inherited genetic factors, many environmental, lifestyle and other factors could have important and complicated impacts on a patient's response to a drug.

Researchers will continue to use clinical studies to link genetic and biomarker information to patient responses. As such information becomes part of patient records, it will also become available for use as observational data in uncontrolled studies.

4.2 Estimation

Analyzing large-scale observational data presents several estimation issues, which can be grouped into those related to "high dimensionality" and those associated with the "uncontrolled" nature of the data. Further complicating analysis is that treatments may have heterogeneous effects on different patients, and that this heterogeneity may differ among treatments. In this subsection, we first discuss issues related to high dimensionality, and then issues caused by the uncontrolled nature of the data. Finally, we discuss the challenges of addressing these two types of issues simultaneously in treatment effect analysis. We shall note that the issues and potential solutions we point out below are by no means exhaustive. While they represent some key challenges, the area offers a great potential for many more research opportunities.

4.2.1 High Dimensionality-Related Issues

A direct challenge of working with a large number of variables is that it is unclear which variables are important and how they affect the response variable. Finding the model, which could include nonlinear terms and interactions between variables, with the highest explanatory power is both computationally complex and statistically challenging. The most important statistical challenges are the issues of over-fitting and multi-collinearity, and the problem of multiple testing.

Over-fitting is a common problem in statistical analysis of high-dimensional data, especially when the number of observations is small relative to the number of variables. To illustrate this point, consider a simple example with two observations and one variable. It is possible to perfectly fit the data with a straight-line model that has R-squared equal to one and mean-squared error equal to zero. As such, it has perfect in-sample performance, but will probably have very poor out-sample performance. If someone neglects the issue of over-fitting, he/she might erroneously pick such a model for prediction. Coefficient estimates and p-values in over-fitted models are misleading as well, as they often indicate that the effects of almost all variables are not statistically significant.

Multi-collinearity refers to a situation where some independent variables are highly correlated. An increase in the number of variables increases the probability that some variables are multi-collinear. This is partly because some variables in observational data, though labeled differently, describe similar or closely related attributes (e.g., height and weight) or the same attribute measured at different times (e.g., blood pressure measured throughout the hospital stay). When some variables suffer from multi-collinearity, the estimated effects of these variables will likely have large standard errors, which makes it appear that none of them is statistically relevant.

A systematic approach to addressing both over-fitting and multi-collinearity issues is stepwise variable selection. Forward stepwise selection starts with a null model (i.e., no variable) and recursively adds the remaining variable that has the highest predictive power as measured by R-squared or residual sum of square. Inverse to this, backward stepwise selection starts with the full model (i.e., with all variables) and recursively removes one existing variable that has the lowest predictive power. These stepwise selection processes result in a series of candidate models, whose performances are then compared for model selection. Note, however, this sequential variable selection is often order dependent and computationally inefficient, especially when the number of candidate variables is large.

A more sophisticated approach to addressing over-fitting and multi-collinearity is regularization, which introduces a regularization factor to penalize the number of independent variables that enter the model.

For example, the Least Absolute Shrinkage and Selection Operator (LASSO) is a widely used regularization method that penalizes the number of variables in a regression model and shrinks the coefficients of insignificant or small-impact variables towards zero (Bastani and Bayati, 2018, Bertsimas et al., 2016, Friedman et al., 2001). Under mild conditions, LASSO generates sparse models that identify the truly relevant variables. In contrast with stepwise variable selection, regularization methods can be applied even when the number of variables is larger than the number of observations. However, while methodologically intuitive, regularization methods raise several new questions, including: What is an appropriate amount of regularization? How to compare and select from different models that have similar explanatory powers?

These questions can be answered with cross validation, a technique commonly used in the machine learning literature to compare the performance of different models (Friedman et al., 2001). It uses one set of data (called the "training sample" or "in-sample") to train a model and an independent set of data (called the "testing sample" or "out-sample") to test the model performance, which is usually measured by mean-squared error (MSE) and coverage rate. Cross validation selects a model with the right number and type of variables that best predict outcomes in the testing sample. However, because (1) different training and testing samples might result in different models and (2) there is a potential correlation between training and testing samples due to randomness, the model selection process is usually undertaken using multiple training and testing samples. A popular approach is k-fold cross validation, where the data is randomly divided into k parts. Each time, one of the k parts is held-out for testing and the remaining k-1 parts are used for training. Out-sample MSE of a model is calculated as the average of the k mean-squared errors.

Besides cross validation, researchers have frequently used Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model selection as well. Both AIC and BIC balance the goodness-of-t of the model and its complexity, while BIC penalizes complexity more strongly than AIC and thus leads to sparser models. If the objective is to make better predictions, AIC and cross validations are commonly used. If the objective is to select variables that are truly economically relevant, BIC is typically preferred (Wang et al., 2007).

Finally, multiple testing arises when one uses the same data to test a number of hypotheses. This issue is particularly common in health care settings, because multiple treatments are usually evaluated across multiple metrics for multiple patient groups at multiple providers. For purposes of illustration, suppose the treatment and control groups are patients who received mitral valve repair and replacement, respectively. To guide future treatment decisions, we want to know whether mitral valve repair is superior to replacement for some patients and whether the outcome differences between providers are heterogeneous across different patients. We could partition patients into groups based on their gender and compare repair with replacement as well as different providers for each group in terms of mortality, complication, readmission, graft failure, etc. The more quality metrics used for comparison, the more likely that the treatment and control groups differ on at least one quality metric due to random errors. Similarly, the more patient groups and providers we consider, the more likely that mitral valve repair will be found to be superior for at least one patient group at one provider.

To address the issue of multiple testing, one needs to adjust the significance levels used for rejecting a null hypothesis. Bonferroni and Benjamini-Hochberg are two commonly used approaches partly due to their simplicity (see Bonferroni (1936), Benjamini and Hochberg (1995) for details). Note these tests rely on the assumption that different hypotheses are independent of each other, which might not be true in health care settings. For example, a patient who is more likely to have a complication is also more likely to have a readmission. A better and more complicated approach to finding adjusted significance levels is simulation. The null hypothesis is that the treatment does not have any effect on any observation for any metric. We can then randomly permute the treatment dummies among observations and test all

hypotheses to obtain a set of p values. After that, we repeat the permutation N times and count the number of times at least one test is deemed significant based on the target significance level a. The adjusted significance level can be calculated as a/N.

4.2.2 Uncontrolled Data-Related Issues

The uncontrolled nature of observational data introduces a number of challenges beyond those presented by the high-dimensionality of the data. In randomized controlled experiments, because observations are randomly assigned to treatment and control groups, we can control the number of observations in each group and the treatment effect can be estimated using the average outcome difference between the treatment and control groups. In observational studies, because we cannot randomly assign treatments to individuals, the treatment group may have systematically different attributes than the control group, which may result in estimation biases. The sample size of the treatment may also be insufficient to power the analysis, because unlike randomized experiments, the sample size of the treatment group in an observational study is not chosen by researchers for the purpose of detecting a given effect at the desired level of significance. In addition to bias and sample size issues, censoring is common in health care observational data, because an event (e.g., mortality) may occur beyond the data collection period. Below we first review a list of common sources of biases, which include omitted variable, sample selection and missing value, and then discuss issues of sample size and censored data.

Omitted variable bias occurs because observational data may not include all variables of interest. While it is safe to omit a variable that is pure noise, omitting a variable that directly affects the outcome not only reduces the prediction accuracy of a model, but also creates biases in estimating the coefficients of all other variables that correlate with the omitted variable. For example, health-conscious patients may be more likely to seek treatments at a Center of Excellence (CoE). At the same time, these patients are healthier due to their lifestyles. However, we may not be able to observe from observational data whether a patient is health-conscious or not. Simply ignoring this unobservable variable will bias the estimate of the quality gap between CoEs and non-CoEs.

Sample selection bias occurs when patients described by observational data are not a representative sample of the population of interest. For example, some transplant centers decide whether to admit a patient to their waiting lists based on the acuity of the patient. A very sick patient may be denied access to waiting lists. Therefore, comparing centers based on samples of admitted patients can lead to biased conclusions, especially when different centers have different patient admission practices. Similarly, selection bias is the primary threat to using observational data in comparative treatment analysis. For example, early-stage cancer patients with better prognoses are more likely to receive milder forms of chemotherapy with fewer side effects, while advanced-stage cancer patients with poorer prognoses are more likely to receive stronger drugs with harsher side effects. Hence, both outcomes and side effects are strongly influenced by the mix of patients treated in a center. Giordano et al. (2008) demonstrated how selection bias can lead to different or even opposite conclusions to those from randomized clinical trials in cancer research.

Missing value bias is common in observational data that are collected over a long period of time and from different sources (e.g., multiple hospitals). If only a small portion of the values are missing, and the missing values happened completely at random, we can safely delete these observations without biasing the results. However, if values are missing in a systematic way or for a specific group of observations (e.g., patients with certain comorbidities), simply ignoring the observations will create estimation biases.

Popular techniques for addressing these biases include instrumental variable, panel data, difference-in-differences, regression discontinuity, selection models, and others. The instrumental variable (IV) method explores the variation of an exogenous variable that correlates with the endogenous variable but does not

directly correlate with the outcome variable. The IV approach extracts and uses only the exogenous variation of the variable of interest in the estimation, and obtains an unbiased estimate of its impact (see e.g., Bartel, Chan and Kim, 2016, Chan et al., 2016, Freeman et al., 2016, Ho et al., 2000, KC and Terwiesch, 2011, KC and Terwiesch, 2012, Kim et al., 2014, Lee et al., 2017, Lu et al., 2017, McClellan et al., 1994, Xu et al., 2017). For example, Hadley et al. (2010) showed that a properly chosen instrumental variable can correct the potential selection bias of observational data and provide consistent results as in randomized trials in the context of prostate cancer. Panel data models exploit longitudinal variations to take out unobserved time-invariant individual fixed effects (see e.g., Clark and Huckman, 2012, Clark, Huckman and Staats, 2013, KC, 2013). The difference-in-differences method relies on the assumption that the treatment and control groups would have had parallel trends if there had been no treatment (see e.g., Song, Tucker and Murrell, 2015). It uses group fixed effects to capture time-invariant differences between the two groups and time fixed effects to capture the trends. Regression discontinuity exploits sharp changes in the variable of interest and compares outcomes of observations lying slightly above and below the threshold (see e.g., KC, 2018). Other techniques involve direct modeling of the bias generating process. For example, the Heckman Two-Stage selection model (see Heckman (1976, 1977) for more details) corrects for biases caused by sample selection or non-random missing data. Which technique is most appropriate depends on the context and the availability of exogenous variables.

Even with big observational data, insufficient sample size may still be an issue, particularly as the stratifications in stratified medicine become finer. In experimental studies, researchers use power analysis to determine the sample size required to detect a certain treatment effect at a given significance level, sample size is a critical part of the experiment design. However, in observational studies, because researchers have no control over the number of patients who receive a given treatment, the treatment group may have a very small number of observations, which creates estimation difficulties. For example, suppose we want to compare the mortality rates of two hospitals for a particular surgical procedure. One hospital treated 200 patients with 4 deaths and the other hospital treated 10 patients with no deaths. Which hospital is better? Obviously, the second hospital has a lower mortality rate, but we cannot conclude that it is better because the small sample size makes it likely that this rate was largely due to randomness.

The sample size issue described above is usually difficult to address without borrowing additional information from other sources. If we can construct a prior distribution regarding the outcome based on outside information, an empirical Bayes method can be used to estimate the effect of a treatment or compare multiple treatments. In the surgical example described above, such a prior could be the overall distribution of mortality rates across all hospitals, the relationship between a hospital's volume and its outcome, or the correlations between outcomes of similar procedures. Dimick et al. (2009) applied the empirical Bayes method that first estimates the effects of volume on outcomes from the data and then uses updates based on actual performance to predict performance of a specific hospital. Similar ideas might be suited to analyzing the effectiveness of more advanced chemotherapy treatment for patients of uncommon cancer types with a given chromosome gain or loss (e.g., a Wilms tumor patient with 16q chromosome loss). For example, one could construct a prior using data from patients of similar cancer types with the same chromosome change (e.g., all other kidney cancer patients with 16q loss) and then use the Bayes method to update this prior specifically for Wilms tumor patients.

Finally, censoring refers to the coding of a continuous variable into one number when the variable is larger or smaller than a specific value. There are three types of censoring: left censoring, interval censoring and right censoring, which correspond to different positions of the censored value. Right censoring is the most common in survival analysis, as an event of interest might happen beyond the study period. For example, suppose we use survival time as a quality metric to compare different hospitals for liver transplant surgery. The study period is all the years when data were collected, and the variable of interest is length of survival after transplant. However, we can only observe how long a patient survived only if the patient died before the last date of data collection.

One approach for dealing with censored data is to convert the variable of interest into binary indicators. For example, instead of analyzing how long a patient survived, we could analyze whether a patient survived more than one year after the transplant. However, the choice of the cut-off points may be arbitrary (e.g., one month vs. one year). Moreover, it usually fails to capture the long-term effect of the treatment. Another approach is survival analysis, which explicitly models survival time as a continuous outcome variable. Parametric survival analysis assumes that the underlying distribution of survival time follows a certain known distribution (e.g., exponential) whereas nonparametric survival analysis such as the Kaplan Meier method (see Wooldridge (2010) for more details) focuses on calculating and graphing survival probability as a function of time. A popular method for survival analysis is the Cox proportional hazards model, which is a semiparametric model that can be used to compare the survival time of two or more patient groups of interest.

4.2.3 Issues Related to Treatment Effect Analysis

To use big observational data to inform health care decisions we must simultaneously address issues related to both high-dimensionality and uncontrolled data in the context of treatment effect analysis. Estimating treatment effects is a different objective from predicting outcomes. Even though a more precise estimate of the treatment effect will improve the accuracy of outcome prediction, the model with the most predictive power does not necessarily capture the true treatment effect unbiasedly.

Classic machine learning methods referenced above are proven to be effective for predicting outcomes while addressing the high-dimensionality challenge. By themselves, these techniques are helpful in answering questions such as: How long will a patient with newly diagnosed stage III breast cancer survive? What is the probability that a patient with hypertension (pre-existing) may experience a major complication after a mitral valve surgery? Outcome prediction is particularly useful if a patient has already decided which health care path to follow or which treatment to receive. However, it is not directly useful for a patient who is trying to choose among alternative treatments, unless the causal effects of the treatments have been properly estimated. For example, a patient may ask questions like: How is my quality of life likely to differ if I receive a kidney transplant instead of continuing renal dialysis? What are the relative risks of complications if I choose to get a stent instead of a bypass graft?

While the effect of a treatment can be calculated by comparing predicted outcomes with and without the treatment, the results may be statistically misleading for two reasons. First, the best model for outcome prediction may not be the best for treatment effect estimation. For example, if age is an important factor affecting survival and race is an important factor affecting the effectiveness of a treatment, a model focusing outcome prediction will include age as a key predictor whereas a model focusing on treatment effect estimation will instead include race as a key predictor. Second, there may be endogeneity issues, which can bias the estimate of treatment effects. Because patients are not randomly assigned to observations, the treatment and control groups may have systematic differences (such as those cause treatment selections) that affect their outcomes. As a result, the effect estimated from a simple subtraction includes not only the true treatment effect but also the systematic difference between the treatment and control groups.

To modify classic machine learning methods for treatment effect analysis, we need to address two main issues. First, because we do not directly observe the treatment effect from data, we cannot use them as a dependent variable to train a model. If we instead use outcomes to train a model, an important variable that affects treatment effect but not outcome may be excluded during variable selection, whereas a less important variable that affects outcome but not treatment effect may be included in the final model. Second, because we do not know the treatment effects for observations in the testing sample, we cannot analyze the performance of a model by calculating its mean squared error or coverage rate or use cross

validation for model selection. If we use outcomes in cross validation, we are likely to choose a model with the best outcome instead of the best treatment effect prediction.

One approach to addressing these issues is to use different penalization factors in LASSO to differentiate variables that affect outcomes from those that affect treatment effects (Imai and Ratkovic, 2013). This approach allows for the possibility that some variables have a relatively small impact on outcomes but a large impact on treatment effects. This approach, however, requires institutional knowledge to distinguish between the two types of variables. An alternative approach first transforms the original outcomes and then applies the standard LASSO with the transformed outcomes and original independent variables for treatment effect analyses (Signovitch, 2007).

These approaches identify variables that significantly affect treatment effects when there are no pressing concerns of endogeneity. To address potential endogeneity issues associated with observational data, we need to integrate econometric methods such as the instrumental variable into machine learning models to obtain unbiased estimates of the treatment effects. For example, Li et al. (2017) combined LASSO and the instrumental variable technique to identify price competition in high-dimensional space. Approaches like this offer the best of both worlds by using machine learning to deal with big data and econometrics to establish causality.

Existing studies that analyze the causal effect of a treatment have focused on the average treatment effect. These studies implicitly assume that a treatment has the same effect for all observations. However, it is possible that the same treatment has a positive effect on some patients but a negative or no effect on others. In cases when a treatment has a positive (or negative) effect on all patients, it is possible that the magnitude of the effect differs by patient. Recognizing the differences in patients' responses to the same treatment, researchers have called for stratified medicine that identifies patient groups with heterogeneous treatment effects.

One approach to identifying patient groups is to first interact the treatment dummy with patient characteristics and then include both the interaction terms and other variables in a regression model such as LASSO for variable selection. The main difficulty with this approach is that the number of interaction terms increases exponentially with the number of variables. As a result, this approach is computationally expensive and requires a large number of instruments (or other econometric techniques) to obtain unbiased estimates of all interaction terms.

An alternative approach is to partition observations into groups for subgroup analyses. The first challenge with this approach is that the number of ways to partition patients increases exponentially with the number of patient characteristics. It is not clear how many patient groups we should have and which characteristics should be used for partitioning. The second challenge is that there is a clear tradeoff between relevance and reliability. A finer partitioning of patients provides more relevant information but may not have enough statistical power due to a reduced sample size. A coarser partitioning of patients will have a larger sample size, but the information provided may not be as relevant.

These challenges can be addressed by tree-based methods, which are data-driven automatic processes that partition observations into groups such that observations in the same group have similar treatment effects and those across different groups have different treatment effects. A tree method usually starts at the root where all observations are in the same group (called "parent" node) and recursively partitions observations into "child" nodes using the variable that increases in-sample goodness of fit the most. It then treats each child node as a parent node and continues the partitioning until a stopping criterion is reached. Finally, it uses cross validation to select the tree that has the best out-sample performance.

While it is straightforward to calculate the treatment effect for a group in randomized controlled trials using average outcome differences between the treatment and control groups (Athey and Imbens, 2016), we need to address potential endogeneity issues with observational data and use the instrumental variable or another econometric method to obtain unbiased estimates of treatment effects. Furthermore, most existing tree-based methods are myopic in partitioning observations and may not achieve the best overall prediction accuracy. Dimitris et al. (2017) proposed an optimal tree to address this issue, but this approach is computationally expensive if the dimension of the problem is high.

Finally, the main challenge of using big observational data to inform medical decisions at the individual patient level is that we cannot observe both the treatment and no-treatment conditions for a patient and no two patients are exactly alike. Even if we are able to perfectly match two patients based on observable characteristics, there is no guarantee that the two patients have the same unobservable characteristics. To address this main challenge and move practice closer to precision medicine, we can look to the inevitable increase in the amount of health care data and rise of real-time monitoring through use of wearable devices. In many health care settings, it is reasonable to assume that only a finite number of variables affect treatment effects, which means an increase in the number of observations will enable us to partition patients into finer groups without losing statistical power. Real-time monitoring by wearable devices allows us to collect observations for the same patient on a daily or even hourly basis. If a treatment is assigned to the same patient at random times or based on observable variables, these observations constitute trials for the same patient (see e.g., Klasnja et al., 2015). For example, a wearable device might remind the wearer to take deep breaths as an anxiety reducing therapy (see e.g., Sarker et al., 2017). With some randomization of the reminders and measurement of the physical responses, the device could optimize a breathing strategy for a given individual (see e.g., Walton et al., 2018). Finally, if the previously mentioned privacy concerns can be addressed, data from such devices could be pooled and used to determine the individual characteristics that make people most and least receptive to breathing therapy.

In summary, heterogenous treatment analysis with high-dimensional observational data offers a wealth of research opportunities. New research in this area is emerging (see, e.g., Athey et al., 2017, Wang, Li and Hopp, 2017a & 2017b, Boruvka et al., 2017) and has the potential to sharpen precision medicine protocols for a vast range of patients.

4.3 Modeling and Optimization

Once we have used big data to estimate patient responses to various health care alternatives, the problem becomes how to use the results to facilitate better decisions across the health care system. These decisions include choice of type of health care and specific providers by patients, selection of specific treatments by patients and providers, strategic positioning and process improvement decisions by providers, reimbursement and incentive structure decisions by payers, and many others. We list some important decisions that present challenges amenable to the modeling and optimization skills of OM scholars below.

4.3.1 Individual Patient Level

In decentralized health care systems where patients can choose the type of treatment and the specific providers for the treatment, observational data can be used to understand how patients make choices and identify barriers that prevent patients from finding health care most suitable for them. When choosing a health care provider, patients consider not only outcome differences between providers but also travel distance, waiting time, insurance co-pay, etc. Patient choice models help understand not only the relative weights patients place on different factors but also how much outcomes would improve if one or more barriers are removed. For example, Wang et al. (2015) studied the impact of quality information, travel distance and insurance on patients' choice of cardiac surgeons for mitral valve surgeries and found that

lack of quality information is the most important barrier preventing mitral patients from choosing the best care and reducing this barrier could improve mitral valve repair rate by 13%.

Although choice models using observational data can help understand what patients *are* doing, analytical models using OM techniques such as optimization are needed to tell patients and their providers what they *should* do. Compared with traditional models that rely on wide range of assumptions, data-driven analytical models estimate parameters based on observational data and use them as input for modeling and optimization. One application of data-driven analytical models is multi-criteria decision making, where multiple treatments (e.g., surgery, stents and statins for carotid disease) are available and a patient needs to consider multiple factors (e.g., recovery time from treatment, risk of complications and life expectancy) in choosing a treatment. Multi-criteria decision making is challenging because a given treatment may look better on one criterion but worse on others. OM techniques such as multi-objective optimization and the analytical network/hierarchical approach (Saaty, 2013) can help patients choose the best treatment based on their individual preferences. These techniques can also help patients with sensitivity analyses that help them see how their choice depends on outcome estimates and/or their personal preferences.

In addition to multiple criteria, health care decisions often involve multiple periods. For example, patients with chronic diseases are typically faced with a series of decisions to manage their condition. In some cases, such decisions are spread over time because treatment options change over time. For example, patients waiting for an organ transplant must decide whether to accept a currently available organ or to wait for a future and potentially better organ. Metastatic cancer patients can face similar decisions involving uncertain future options as they wait for improved chemotherapy options to become available. This type of multi-period decision making is challenging, because a patient's health status evolves over time. In the case of organ transplant, if a patient decides to reject an organ and stay on the waiting list, he/she might get a better offer in the future, but his/her health state might get worse. Problems like these can be addressed using familiar methods, such as linear programming and dynamic programming (Alagoz et al., 2009), but may also require new data analytics methods to update evolving options, risks, and patient characteristics to properly parameterize the models.

4.3.2 Health Care System Level

As data becomes more and more transparent, health care providers of all kinds will be faced with decisions of what services to offer. If patients can see which providers are substandard for which procedures, they will be able to selectively avoid them. Since hospitals will no longer be able to use the halo from their strengths to hide their weaknesses, it will become increasingly difficult to offer a full range of medical services. This will present hospitals and other medical providers with "invest or specialize" scenarios. They will either need to invest in process improvement to make their weakest services competitive or eliminate those services and compete on their strong services. Such decisions will be complicated by the fact that there are synergies between services (e.g., a strong transplant program for one type of organ may offer infrastructure and marketing advantages with which to build a transplant program for another type of organ). These complex strategic planning problems will require modeling support, which will need grounding in the types of data-driven evaluation of performance we have discussed in this paper.

Observational data can also help health systems improve outcomes through better matching of patients to facility and/or provider. As described earlier, the effect of a treatment may be heterogeneous for different patients and such heterogeneity may differ across treatments. Here, the "treatment" could be the actual medical treatment (e.g., surgery, stent, statin), but it could also be the type of facility (e.g., doctor's office, urgent care clinic, community hospital, research hospital) or the individual provider (e.g., specific surgeon). Heterogeneous outcomes imply that a given facility or provider may be better suited to one

type of patient than another and that the "best" facility or provider will be different for different patients. Using big data analysis to uncover such heterogeneity in outcomes can enable health systems to better guide patients (e.g., through physician referrals) to the types and sources of care that best meet their individual needs. By taking into account both clinical quality and operational efficiency, improvements in the matching process can help systems provide better health at a lower cost.

Modeling and optimization can also help health systems respond to pay-for-performance systems. Increased transparency of hospital performance data, made possible by more sophisticated big data analytics, will allow payers such as CMS to tie reimbursements more closely to the actual value hospitals provide to patients. An example of early efforts to do this are the hospital readmission reduction program (HRRP), which penalizes hospitals with excessive 30-day readmission rates and the hospital acquired condition reduction program (HACRP), which penalizes the worst quartile of hospitals with regard to hospital acquired infections. Because these programs base penalties on coarse evaluations of under-performing hospitals, they create incentives that may not align with the goal of promoting improvements in hospital performance. For example, Zhang et al. (2016) developed a game-theoretic model that captures the competition among hospitals induced by HRRP's benchmarking mechanism and found that low-performing hospitals prefer paying penalties to reducing readmissions. Better performance evaluation will create opportunities for more targeted and more effective pay-for-performance mechanisms, which will require modeling and optimization support to exploit.

4.3.3 Societal Level

If a health care system were governed by a central planner who decides how to allocate limited treatment resources to different patients, heterogeneous and personalized data could be used directly to make better patient assignment decisions. But even this highly simplified version of the health care resource allocation problem is not entirely straightforward. While it is generally simple to solve an allocation problem in which both treatment effects and disease states are deterministic, the problem becomes much harder if we consider uncertainty in the effect of a treatment and the progression of a disease. To illustrate this point, consider two treatment types (Treatment 1 and 2) and two disease states (State 1 and 2). Treatment 1 is more effective and expensive than Treatment 2, but none of the treatments can cure the disease. Furthermore, the disease state evolves over time as a function of prior state and treatment type. Because budget is limited in each period, the central planner must decide which patients to treat and the type of treatments to provide. This type of problem can be formulated using linear programming or dynamic program but is generally difficult to solve when the number of periods, treatment types or disease states under consideration is large. Zayas-Caban et al. (2017) proposed an asymptotically optimal heuristic based on a linear programming relaxation to the original stochastic formulation.

Since no country or health care system operates in pure central planning mode, analyses of such systems are used as guides for narrow decisions that do involve central planning, such as which procedures should be covered in medical plans, which patients should be prioritized in allocating scarce resources, and what guidelines and regulations should be adopted. Specific examples include allocating organs for transplant (Ata, Freidewald, Randa 2017), optimizing colonoscopy screening for colorectal cancer prevention and surveillance (Erenay, Alagoz, Said 2014)), allocating scarce healthcare resources in developing countries (Griffin, Keskinocak, Swann 2013), optimizing breast screening policies given heterogeneous adherence (Turgay et al. 2016).

All real-world health systems are at least partially decentralized because patients have at least some choice about what treatment to get and where to get it. Such systems still require modeling and optimization to design and plan. For example, to help system managers evaluate the value of a new type of information (e.g., EMR information made possible by Medicalchain), we need to be able to predict how the new information will affect patient decision making and how this will impact patients and other

stakeholders. An example of such a model is that of Wang et al. (2017), which analyzed the relative value of population-average and patient-specific quality information about cardiac surgeons. This involved modeling patient choice of surgeon as a function of outcome quality, travel distance and waiting time on the surgical schedule and combining this with a queueing model to estimate patient waiting time. The results suggested that societal benefits (i.e., sum of patient utility) from using patient-specific information about mitral valve surgery outcomes are comparable to those achievable by enabling the best surgeons to treat 10%-20% more patients under population-average information. Analyses like these can help identify areas where more detailed information can facilitate precision medicine to achieve the greatest societal benefits.

Modeling and optimization can also help understand and improve interactions between system managers and patients that govern how resources are used. For example, in organ transplant systems, UNOS plays the role of system manager responsible for organ allocation. When an organ becomes available, patients with compatible blood type and antibody-antigen are sequenced according to their sickness, waiting time, proximity to donor, etc. The organ is then offered to patients in this sequence. UNOS quickly realized that patients who were more likely to receive an allocation were also more likely to reject the allocation. Hence, by using historical data to understand patients' decisions, UNOS developed the Liver Simulation Allocation Model to compare potential outcomes of alternative allocation policies. Several studies have made use of this simulation to incorporate patient choice into optimization models that improve the allocation process (Su and Zenios, 2004, 2005 & 2006, Akan et al, 2012). As more performance data becomes available as a result of the combined big data and precision medicine revolutions, analogous analyses of other scarce resources (e.g., elite surgeon time, experimental drugs, new imaging technologies, etc.) will become possible. POM scholars are well-equipped to provide the needed analytic innovations.

4.4 Implementation

Achieving precision medicine, as we have described it here, will require more than solving statistical and analytics problems or creating simulation and optimization models. We must also find ways to communicate complex information about treatment options and outcomes to patients and providers in ways they can use it. This will require better displays (e.g., customized rankings in web pages) and tools for dealing with multi-criteria decisions (e.g., aids to help users figure out weights for criteria or mechanisms for using partial weight information to choose between alternatives).

A number of observers have noted that public reporting of health care outcomes has been less effective than anticipated in altering patient and provider behavior (see e.g., DeVore et al. 2012, Ryan et al. 2012 and Smith et al. 2012). Saghafian and Hopp (2018) concluded that the primary reason for this is that patients have acted as though they either do not have or do not understand the publically reported data. This implies that information contained in this data has not been communicated in a clear and usable way. If this is indeed the case, then there is considerable opportunity to advance progress toward precision medicine by better communicating the information being generated by the big data revolution. POM scholars can contribute to research into how patients perceive information about uncertain outcomes, how trust (e.g., of doctors, insurance companies, government) influences patient reception of information from various sources, and many other practical questions related to the use of information in real-world settings.

However, having and understanding the data needed to make effective decisions for precision health care is not enough. Both patients and providers must act upon it. If other factors distort treatment decisions, patient and societal benefits will fall short of their potential. Three issues that may prevent a provider from acting in patients' best interests are: (1) Some hospitals are profit-driven, so they may pick patients or prescribe medications based on how much revenue they could generate rather than how well the

patients could be treated; (2) Hospitals that are rated and reimbursed according to imperfectly risk-adjusted outcomes may intentionally avoid very sick patients who are likely to have bad outcomes (Dranove et al., 2003); (3) Referral decisions by physicians may be influenced by non-clinical factors such as hospital affiliation, personal relationship, waiting time, etc (Lu and Lu, 2016). Analyses of the impact of big data on precision medicine, as well as designs of pay-for-performance systems, must take behaviors like this into account.

Finally, there are several important roles payers can play in the evolution of precision medicine. First, payers can help guide patients to the most suitable providers by structuring networks, co-pays and other policies to favor effective choices. Second, by combining the above patient incentives with reimbursements that favor high quality outcomes, payers can influence providers to focus on patients and procedures that fit their relative strengths. Third, payers can improve health care through pay-for-performance that incent providers to invest in process improvements that result in better patient outcomes. The ability of payers to carry these out will depend critically on the analytics that underlie performance evaluation.

5. Conclusion

As depicted in Figure 3, the improved data collection and analysis that will result from the trends and activities discussed here will lead to stratified medicine that is increasingly better targeted at individual patient needs. Patients will use the improved information to make more informed decisions about their own health care, while providers will use it to make better strategic planning and process improvement decisions, and payers will use it to design better incentives for both patients and providers. The combined effect will be improved patient outcomes and greater efficiency due to elimination of waste and errors caused by trial-and-error medicine.
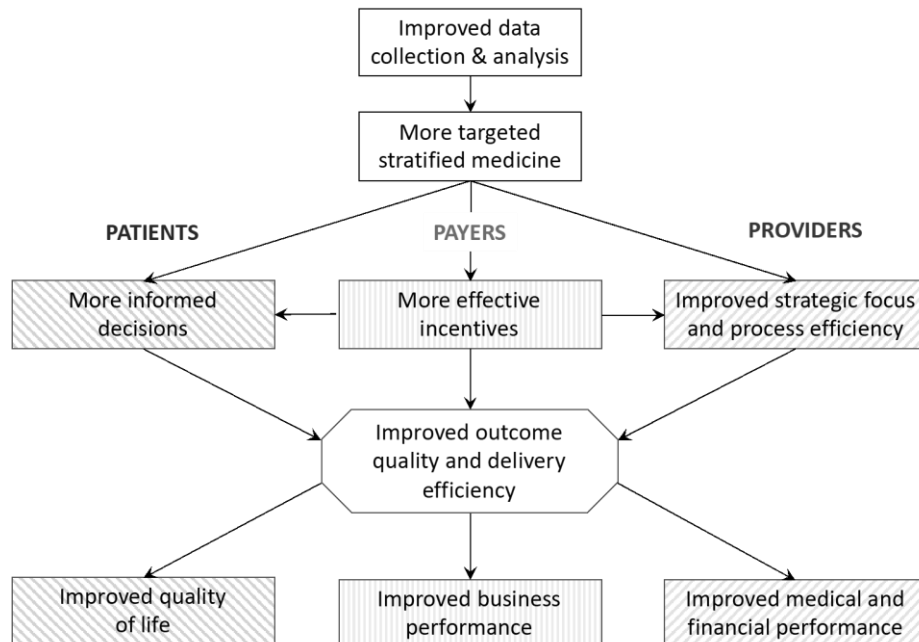
Figure 3: Impact of Precision Medicine on Health Care Stakeholders.

For patients, the primary benefit of better decisions at all levels of the health care hierarchy will be better health. By providing evidence-based guidance to treatment options and provider choices that best suit an individual patient's needs, precision medicine will lead to significant improvements in patient outcomes. Big data will amplify this impact by providing increasingly detailed evidence that increases the power and specificity of precision medicine.

For example, a patient who suffers from severe depression today can expect it to take months, or even years, of consultation with psychiatrists, psychologists, therapists, yoga instructors, etc., to find a combination of medications and lifestyle changes that bring relief. In the coming world of precision medicine, the patient will receive treatments aligned with his/her individual characteristics and will see improvement much faster and will ultimately find a better health outcome.

But patients will benefit from more than better health because of precision medicine. By shifting the basis for medical decisions away from subjective judgement and toward statistical evidence, big data will move the locus of decision-making toward patients. Under trial-and-error medicine, in which providers choose a course of action largely on their personal experience, a patient has little choice but to accept a recommendation based on the provider's judgment. It's not easy for a lay person to argue with an experienced medical professional who says, "in my considered opinion you should do X". But under precision medicine, where there is an explicit base of evidence to indicate a course of action, it will become possible for the patient to argue with a recommendation without disparaging the recommender. For example, suppose an orthopedist recommended a total knee replacement based on data showing an 80% chance of significant pain and mobility improvement and only a 5% chance of an outcome worse than the status quo. A particularly risk-averse patient could reasonably argue that the procedure is too risky for him/her. As data becomes more transparent, patients will become increasingly empowered to participate in more decisions regarding his/her health. The precision medicine revolution will ultimately bring about the end of paternalistic medicine.

In addition to improving health outcomes at the individual and societal levels, precision medicine powered by big data will also reduce costs. Brownlee (2008) estimates that between one-fifth and one-third of health care dollars are spent on unnecessary tests, drugs and treatments (and that such overtreatment is responsible for as many as 30,000 deaths per year). By providing a clear evidence base for determining which interventions are clinically worthwhile, big-data-driven precision medicine will help patients and providers avoid unnecessary treatments and their associated costs. Furthermore, by forgoing treatments that are not well-suited to the patient, precision medicine will lead to fewer errors, complications, and follow-up corrections, all of which will reduce costs. Finally, precision medicine can and should indicate where less invasive measures (e.g., prevention) are the best course of action. In the right situations, these approaches can achieve better health at dramatically reduced costs.

The impact of the precision medicine revolution will impact provider behavior beyond interactions with patients. The detailed patient response data on which precision medicine is based will give clinicians and administrators granular feedback on their strengths and weaknesses. This will allow them to make strategic decisions about their focus, for example choosing which surgical procedures to offer or which patient types to target. It will also help them focus their process improvement efforts on the patients and services where improvements are most needed.

Finally, precision medicine will open up a host of opportunities for payers to sharpen the incentives they provide to both patients and providers, with a goal of achieving better health outcomes at lower costs. When payers have detailed data on the best care options and choice of provider for a given patient, they can design pricing and co-payment schemes that incentivize patients to choose the most cost-effective alternatives. Similarly, they can design pay-for-performance reimbursement schemes that incentivize

payers to seek out patients where they have a comparative quality advantage and to invest in targeted process improvements.

This impressive list of benefits will only be possible if we can resolve the data collection, estimation, modeling and optimization, and implementation issues discussed above. Since these are all challenges POM scholars are well-suited to address, the combined big data/precision medicine revolution is an area where our field can change the world for the better.

Acknowledgements

References

Akan, M., Alagoz, O., Ata, B., Erenay, F. S., Said, A. 2012. A broader view of designing the liver allocation system. *Operations Research* **60**(4) 757-770.

Alagoz, O., Schaefer, A. J., Roberts, M. S. 2009. Optimizing organ allocation and acceptance. *Handbook of Optimization in Medicine.* Springer, Boston, MA, 1-24).

Ata, B., Freidewald, J. J., Randa, A. C. 2017. Organ transplantation. Working paper, University of Chicago, Chicago, IL.

Athey, S., Imbens, G. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* **113**(27) 7353-7360.

Athey, S., Tibshirani, J., Wager, S. 2017. Generalized random forests. Working paper, Graduate School of Business, Stanford University, Stanford, CA.

Ayer, T., Alagoz, O., Stout, N., Burnside, E. 2016. Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science* **62**(5) 1339-1362

Azaria, A., Ekblaw, A., Vieira, T., Lippman, A. 2016, MedRec: Using blockchain for medical data access and permission management. *International Conference on Open and Big Data.* IEEE, 25-30.

Bartel, A. P., Chan, C. W., Kim, H. 2016. Should hospitals keep their patients longer? The role of inpatient and outpatient care in reducing readmissions. Working paper, Columbia Business School, New York, NY.

Bastani, H., Bayati, M. 2018. Online decision-making with high-dimensional covariates. Working paper, Graduate School of Business, Stanford University, Stanford, CA.

Benjamini, Y., Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* **57**(1) 289-300.

Bertsimas, D., Dunn, J. 2017. Optimal classification trees. *Machine Learning* **106**(7) 1039-1082.

Bertsimas, D., O'Hair, A., Relyea, S., Silberholz, J. 2016. An analytics approach to designing combination chemotherapy regimens for cancer. *Management Science*, **62**(5), 1511-1531.

Bonferroni C. 1936. *Teoria Statistica Delle Classi e Calcolo Delle Probabilita*, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze (Libreria Internazionale Seeber, Florence, Italy), Vol 8, pp 3–62.

Boruvka, A., Almirall, D., Witkiewitz, K. Murphy, S.A. 2017. Assessing time-varying causal effect moderation in mobile health. *The Journal of the American Statistical Association.* Forthcoming.

Brownra, S. 2008. *Overtreated: Why too much medicine is making us sicker and poorer.* Bloomsbury, New York.

Bühlmann, P., Van De Geer, S. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Chan, C. W., Farias, V. F., & Escobar, G. J. 2016. The impact of delays on service times in the intensive care unit. *Management Science* **63**(7) 2049-2072.

Clark, J. R., Huckman, R. S. 2012. Broadening focus: Spillovers, complementarities, and specialization in the hospital industry *Management Science* **58**(4) 708-722.

Clark, J. R., Huckman, R. S., Staats, B. R. 2013. Learning from customers: Individual and organizational effects in outsourced radiological services. *Organization Science* **24**(5) 1539-1557.

DeVore, A.D., B.G. Hammill, N.C. Hardy, Z.J. Eapen, E.D. Peterson, A.F. Hernandez. 2012. Has public reporting of hospital readmission rates affected patient outcomes? *Journal of the American College of Cardiology* **67**(8) 570-577.

Dimick, J. B., Staiger, D. O., Baser, O., Birkmeyer, J. D. 2009. Composite measures for predicting surgical mortality in the hospital. *Health Affairs* **28**(4) 1189-1198.

Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M. 2003. Is more information better? The effects of "report cards" on health care providers. *Journal of political Economy* **111**(3) 555-588.

Erenay, S., O. Alagoz, A. Said. 2014. Optimizing Colonoscopy Screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management* **16**(3), 381-400.

Federal Drug Administration. 2018. Precision Medicine. https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/PrecisionMedicine-MedicalDevices/default.htm. [Accessed February 23, 2018]

Freeman, M., Savva, N., Scholtes, S. 2016. Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science* **63**(10) 3147-3167.

Friedman, J., Hastie, T., Tibshirani, R. 2001. *The elements of statistical learning, Vol. 1.* Springer Series in Statistics, New York, 337-387.

Giordano, S. H., Kuo, Y.-F., Duan, Z., Hortobagyi, G. N., Freeman, J., Goodwin, J. S. 2008. Limits of observational data in determining outcomes from cancer therapy. *Cancer* **112**(11) 2456-66.

Glance, L. G., Dick, A., Osler, T. M., Li, Y., Mukamel, D. B. 2006. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. *Medical care* **44**(4),311-319.

Griffin, J., Keskinocak, P., Swann, J. 2013. Allocating scarce healthcare resources in developing countries: A case for malaria prevention. B. Denton, ed. *Handbook of Healthcare Operations Management*. Springer, New York, NY, 511-532.

Hadley, J., Yabroff, K. R., Barrett, M. J., Penson, D. F., Saigal, C. S., Potosky, A. L. 2010. Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. *Journal of the National Cancer Institute* **102**(23) 1780-1793.
Heckman, J. J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**(4) 475-492.

Heckman, J. J. 1977. Sample selection bias as a specification error. *Econometrica* **47(1)** 153–161.

Ho, V., Hamilton, B. H., Roos, L. L. 2000. Multiple approaches to assessing the effects of delays for hip fracture patients in the United States and Canada. *Health Services Research* **34**(7) 1499-518.

Imai, K., & Ratkovic, M. 2013. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics* **7**(1) 443-470.

Kc, D. S., Terwiesch, C. 2011. The effects of focus on performance: Evidence from California hospitals. *Management Science* **57**(11) 1897-1912.

Kc, D. S., Terwiesch, C. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* **14**(1) 50-65.

Kc, D. S. 2013. Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management* **16**(2),168-183.

Kc, D. S. 2018. Heuristic thinking in patient care. Working paper, Goizueta Business School, Emory University, Atlanta, GA.

Kim, S. H., Chan, C. W., Olivares, M., Escobar, G. 2014. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* **61**(1) 19-38.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., Murphy, S. A. 2015. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* **34**(S) 1220.

Kruse, C., R. Goswamy, Y. Raval, S. Marawi. 2016. Challenges and opportunities of big data in health care: A systematic review. *JMIR Medical Informatics* **4**(4) e38.

Lam, Y. W. 2013. Scientific challenges and implementation barriers to translation of pharmacogenomics in clinical practice. *ISRN Pharmacology* Vol 2013, Artilce ID 641089.

Lee, D., Li, J., Wang, G., Croome, K., Burns, J., Perry, D., Nguyen, J., Hopp, W., Taner, B. 2017. Looking inward: impact of operative time on patient outcomes. *Surgery* **162**(4) 937-949.

Li, J., Netessine, S., Koulayev, S. 2017. Price to compete… with many: how to identify price competition in high-dimensional space. *Management Science*. Forthcoming.

Li, X., J. Qin. 2017. Anonymizing and sharing medical text records. *Information Systems Research* **28**(2) 332-352.

Little, R. J., Rubin, D. B. 2014. *Statistical analysis with missing data, Vol. 333*. John Wiley & Sons, New York.

Lu, L. X., Lu, S. F. 2016. Distance, quality, or relationship? Interhospital transfer of heart attack patients. *Production and Operations Management*. Forthcoming.

McClellan, M., McNeil, B. J., Newhouse, J. P. 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association* **272**(11) 859-866.

Miller, A., C. Tucker. 2017. Privacy protection, personalized medicine, and genetic testing. *Management Science - Articles in Advance*, 1-22. [https://doi.org/10.1287/mnsc.2017.2858]

National Research Council. 2011. Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. Washington, DC: The National Academies Press. https://doi.org/10.17226/13284.

Rabbi, M., Aung, M. H., Choudhury, T. 2017. Towards health recommendation systems: an approach for providing automated personalized health feedback from mobile data. Rehg, J., Murphy, S., Kumar, S., eds. *Mobile Health.* Springer, Cham, 519-542.

Ryan, A.M., B.K. Nallamothu, J.B. Dimick. 2012. Medicare's public reporting initiative on hospital quality had modest or no impact on mortality from three key conditions. *Health Affairs* **31**(3) 585-592.

Saaty, T. L. 2013. The modern science of multicriteria decision making and its practical applications: The AHP/ANP approach. *Operations Research* **61**(5),1101-1118.

Saghafian, S., Hopp, W. 2017. Can public reporting cure healthcare? The role of quality transparency in improving patient-provider alignment. Working paper. Harvard University, Cambridge, MA.

Sarker, H., Hovsepian, K., Chatterjee, S., Nahum-Shani, I., Murphy, S. A., Spring, B., ... Kumar, S. 2017. From markers to interventions: the case of just-in-time stress intervention. Rehg, J., Murphy, S., Kumar, S., eds. *Mobile Health*. Springer, Cham, 411-433.

Seung-Jung Park, Jung-Min Ahn, Young-Hak Kim, Duk-Woo Park, Sung-Cheol Yun, Jong-Young Lee, Soo-Jin Kang, Seung-Whan Lee, Cheol Whan Lee, Seong-Wook Park, Suk Jung Choo, Cheol Hyun Chung. 2015. Trial of everolimus-eluting stents or bypass surgery for coronary disease. *New England Journal of Medicine* **372** 1204-1212.

Signorovitch, J. E. 2007. Identifying informative biological markers in high-dimensional genomic data and clinical trials. Doctoral dissertation, Harvard University, Cambridge, MA.

Smith, M.A., A. Wright, C. Queram, G.C. Lamb. 2012. Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. *Health Affairs* **31**(3) 570-577.

Song, H., Tucker, A. L., Murrell, K. L. 2015. The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* **61**(12) 3032-3053.

Spigel, D. R. 2010. The value of observational cohort studies for cancer drugs. *Biotechnology Healthcare* **7**(2) 18–24.

Su, X., Zenios, S. 2004. Patient choice in kidney allocation: The role of the queueing discipline. *Manufacturing & Service Operations Management* **6**(4) 280-301.

Su, X., Zenios, S. A. 2005. Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations Research* **53**(3) 443-455.

Su, X., Zenios, S. A. 2006. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science* **52**(11) 1647-1660.

Unger, J. M., Cook, E., Tai, E., Bleyer, A. 2016) Role of clinical trial participation in cancer research: barrier, evidence, and strategies. *American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Meeting* **35** 185–198.

Walton, A., Nahum-Shani, B., Crosby, L., Klasnja, P., Murphy, S. 2018. Optimizing digital integrated care via micro-randomized trials. *Clinical pharmacology and therapeutics*. Forthcoming.

Wang, G., Li, J., Hopp, W. 2017a. An instrumental variable tree approach for detecting heterogeneous treatment effects in healthcare and other observational studies. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.

Wang, G., Li, J., Hopp, W. 2017b. Personalized health care outcome analysis of cardiovascular surgical procedures. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.

Wang, G., Li, J., Hopp, W. J., Fazzalari, F. L., Bolling, S. 2015. Hospital quality and patient choice: An empirical analysis of mitral valve surgery. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.

Wang, G., Li, J., Hopp, W., Fazzalari, F., Bolling, S. 2017. Using patient-specific quality information to unlock hidden health care capabilities. *Manufacturing & Service Operations Management*. Forthcoming.

Wang, H., Li, R., Tsai, C. 2007. Tuning parameter selector for the smoothly clipped absolute deviation method. *Biometrika* **94**(3) 553–568.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, MA.

Wu, D. 2016. Shock spillover and financial response in supply chain networks: Evidence from firm-level data. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.

Xu, Y., Armony, M., Ghose, A. 2016. The effect of online reviews on physician demand: A structural model of patient choice. Working paper, Stern School of Business, New York University, New York, NY.

Zayas-Caban, G., Jasin, S., Wang, G. 2017. An asymptotically optimal heuristic for general non-stationary finite-horizon restless multi-armed multi-action bandits. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.

Zhang, D. J., Gurvich, I., Van Mieghem, J. A., Park, E., Young, R. S., Williams, M. V. 2016. Hospital readmissions reduction program: An economic and operational analysis. *Management Science* **62**(11) 3351-3371.