

Employee Health Plans Powered by Analytics



Wallace Hopp,
Ross School of Business,
University of Michigan

Jun Li,
Ross School of Business,
University of Michigan

Soroush Saghafian,
Harvard Kennedy School,
Harvard University

Guihua Wang, Jindal
School of Management,
University of Texas at Dallas

A promising new business model for the employee health plans of large firms bypasses insurers and instead uses direct contracts with hospitals which have been designated as centers of excellence. Wallace Hopp, Jun Li, Soroush Saghafian, and Guihua Wang describe how combining this model with cutting edge analytics could revolutionize the delivery of high quality, cost-efficient health care.

©shutterstock/janews

Whether they realize it or not, most large firms in America are in the health care business. Approximately 153 million people in the US, nearly half the population, rely on employer-sponsored plans for health coverage. Sixty-one percent of those plans are either partially or completely employer funded. Employers therefore insure 93 million people, which is more than either Medicaid (70 million) or Medicare (40 million). And with average premiums of \$7,188 for single coverage and \$20,576 for family coverage in 2019, health care is the second largest expense in the operating budget of most companies, after wages.¹

Whether they realize it or not, most large firms in America are in the health care business. Approximately 153 million people in the US rely on employer-sponsored plans for health coverage.

Center of Excellence Programs

Rising health care costs have driven up the premiums for employer-sponsored health insurance by 22 percent over the past five years and 54 percent over the past decade. Understandably, firms have responded by devising an array of measures to contain costs.¹ These measures include increased employee premiums and deductibles, consumer-directed health plans, telehealth systems, and many more. Large firms, including Walmart, Lowes, GE, Boeing, PepsiCo, and many others, have also reduced costs by eliminating the intermediary insurance company and contracting directly with *centers of excellence*

(COEs) to provide their employees with major care such as surgeries, transplants, cancer treatment, and other procedures.² High quality medical providers also promote cost efficiency over the long term because they help patients to avoid costly future complications and readmissions.^{3,4} Finally, by using their size to negotiate favorable rates with COEs, firms can further increase their cost efficiency. As a result, even when firms must pay for their employees' travel to a COE, these plans can produce better health outcomes at lower costs than traditional plans that rely exclusively on local health systems.

Still, these COE programs are almost certainly more expensive and less effective than they could be because they overlook the fact that different people respond differently to the same treatment. For example, if the Cleveland Clinic is designated as a COE for cardiac patients, the program might encourage (e.g., by travel and lodging subsidies) or require all patients who need cardiac procedures to be treated at the Cleveland Clinic. But for some patients, that solution is not optimal. A patient with mitral valve disease and hypertension who lives in another part of the country might be treated equally well, or even better, at a local hospital. If so, the COE program is at risk of overpaying and underperforming.

To avoid guiding patients to the wrong hospital, firms need to determine which patients benefit from a COE and which don't.

The Challenge of Evaluating Hospital Performance

To avoid guiding patients to the wrong hospital, firms need to determine which patients benefit from a COE and which don't. The classic randomized controlled trial approach, which is considered the gold standard for medical research, is not an option because it would be impractical, or even unethical, for firms to randomly send employees to different hospitals so they could study the outcomes. Even if such an experiment were possible once, they certainly could not repeat it to monitor changes over time, as the target hospitals evolved and improved. They must therefore rely on *observational data*, drawn from actual medical results.

Unlike data from randomized controlled trials, in which patients are randomly assigned different treatments, observational data are generated by real-world health care decisions. There is no reason to think that patients in real-world settings choose hospitals randomly. For example, an elite hospital may attract a higher-than-average percentage of patients with clinically complex ailments precisely because it has a reputation for excellence in treating such patients. If we ignore this tendency, called selection bias, and compare the elite hospital's treatment results (in terms of mortality, complications, or readmission rates, for example) with those of other hospitals, the elite hospital will appear to be less successful than it actually is because it is treating patients with more severe illnesses.

The most straightforward way to control for selection bias in observational data is to statistically account for differences in the groups being compared.

The most straightforward way to control for selection bias in observational data is to statistically account for differences in the groups being compared. Including variables in the statistical model for each factor that could bias the outcome is called *risk adjustment*. In a comparison of the complication rates of mitral valve surgery patients at Hospital A and Hospital B, we might include in our statistical model patient demographic and clinical variables such as the ratio of male patients to female, the proportion of patients in each age group, their average body mass index, their rates of comorbidities like diabetes and hypertension, and a variety of other factors. We could then use the model to compute an O/E ratio that compares the actual complication rate (labeled O for “outcome”) with the expected complication rate (E), for each hospital. A hospital with an O/E ratio above 1 has a higher complication rate than would be expected of an average hospital, while a hospital with an O/E ratio below 1 has a lower-than-average complication rate. If Hospital A has a lower O/E ratio than Hospital B, we deem it to have better risk-adjusted performance (because lower complication rates are better).

Yet while risk adjustment is intuitively appealing, it is not always effective. If there are influential variables which we haven’t observed, such as a healthy lifestyle, we cannot correct for them in a statistical model and the selection bias will remain. Fortunately, there are statistical tools that compensate for the selection bias of unobserved variables. One such tool is the instrumental variable (IV) approach.⁵ An IV is a variable that influences treatment assignment, such as the patient’s choice of hospital, but does not directly affect the outcome of that treatment. One such variable might be the distance

between the patient’s home and the hospital. The IV approach corrects for selection biases by viewing the treatment variations as similar to those in an experiment (e.g., one that assigns patients to hospitals based on geographic proximity rather than hospital performance).

While risk adjustment is intuitively appealing, it is not always effective.

Accounting for Differences in Individual Patients

Methods for correcting selection bias, such as the IV approach, can help firms to use observational data to identify COEs. Even with their help, however, the firm may still find that its O/E ratios do not make a decision obvious. To understand why, suppose that a firm is considering hospitals for designation as cardiac COEs and wants to consider the risk of methicillin-resistant *Staphylococcus aureus* (MRSA) infection. The firm consults the Leapfrog Group hospital rating website (leapfroggroup.org), which gets its data from hospital surveys. Leapfrog reports that, in 2019, the Michigan Medicine hospital in Ann Arbor had an O/E ratio for MRSA infections of 0.31. Can the firm interpret that O/E ratio to mean that a given patient’s infection risk at Michigan Medicine is just 31 percent of their risk at an average hospital? Can it at least conclude that the patient’s risk is lower at Michigan Medicine than at a hospital with an O/E ratio of 0.5?

Sadly, the answer to both questions is no. The O/E ratio represents only population-average information, such that the mean risk of all patients treated at Michigan Medicine is 31 percent of the mean risk to those same patients if they were treated at an average hospital. It does not describe the relative risk

for a particular patient (or indeed for any patient). Michigan Medicine could have an O/E ratio much lower than 0.31 for men over 80 who have undergone minor orthopedic surgery and a much higher O/E ratio for women under 50 who have undergone a major cardiac surgery. Michigan Medicine could therefore be a legitimate COE for some patients, but an average or substandard choice for others.

So why not simply compute the O/E ratio for each patient type in order to determine who should consider Michigan Medicine a COE? Suppose that sex, race, age, lifestyle, education and secondary medical conditions (comorbidities) might each affect the relative effectiveness of having a cardiac procedure at Michigan Medicine instead of the Cleveland Clinic. Then suppose that we group people into two sexes, six races, six age groups, four lifestyle categories, and four education levels and that we define twenty-four comorbidities. If these traits can occur in any combination, the total number of distinct categories is $2 \text{ sexes} \times 6 \text{ races} \times 6 \text{ age groups} \times 4 \text{ lifestyle categories} \times 4 \text{ education levels} \times 2^4 \text{ comorbidity combinations} = 19 \text{ billion}$.

Since this number is larger than the number of people on the planet, even the observational data of every cardiac procedure that has ever been performed at Michigan Medicine and the Cleveland Clinic will not allow us to populate every possible subgroup, much less with large enough samples to allow statistical comparison. To fit a standard statistical model to data with so many dimensions, we would need more variables to represent the dimensions and their interactions than there are data points. As a result, the model may fit the data perfectly but not reveal any relationships with statistical significance or have any predictive value. Statisticians call this effect *overfit*.

ting. For firms trying to locate COEs for different patient groups, we can just call it unhelpful.

Very recently, researchers have combined tree-based machine learning with instrumental variables, producing tools that can correct for selection bias.

Analytics and Machine Learning to the Rescue

By applying machine learning to data analytics, we can better address this problem of high dimensionality. Unlike statistical models that are created in advance and then fitted to data, machine learning relies on algorithms that use previous data to extract patterns from raw data. Machine learning methods have proliferated in recent years, rapidly increasing the usefulness of observational data in medical decision making. *Tree-based methods* are particularly useful for finding the best hospitals for various groups of patients.⁶ Very recently, researchers have combined tree-based machine learning with instrumental variables, producing tools that can correct for selection biases and use observational data to accurately estimate how choosing one hospital over another will affect the outcomes of different patient groups.⁷

Suppose a firm is using average complication score, which weights different surgical complications by severity, as the outcome metric to choose between Hospitals A and B as a COE for aortic valve repair (AVR).⁸ Further suppose that O/E ratios indicate that Hospital A has a lower average complication score than Hospital B, such that it would be designated

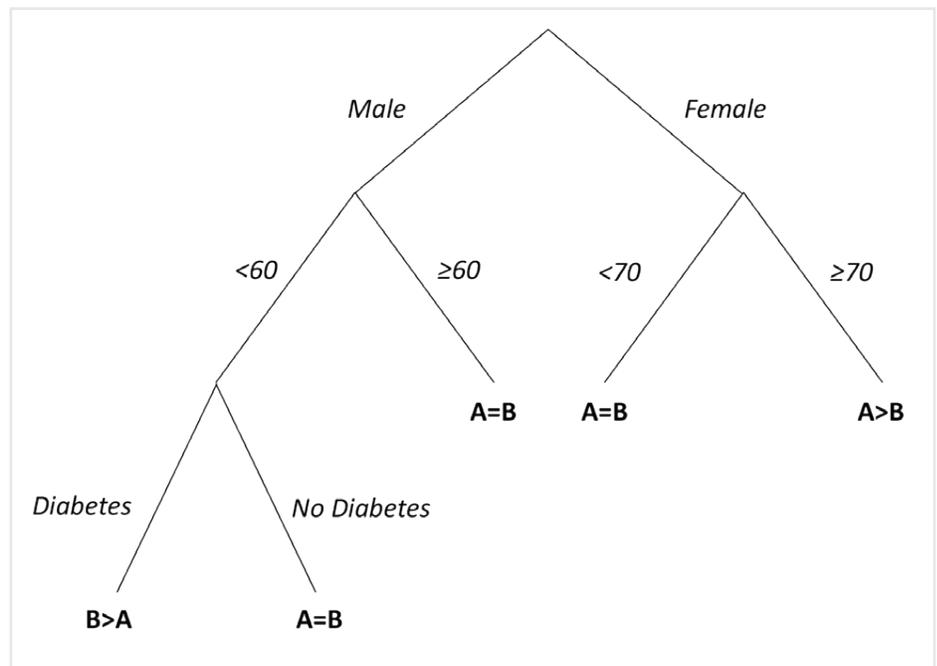


FIGURE 1: Tree-based comparison of Hospitals A and B

the COE under a population-average data comparison. By using data on past procedures at both hospitals, which includes outcomes, patient characteristics, and patient zip code (to estimate their travel distance to each hospital), the firm can use a tree-based algorithm to determine which patient characteristics are more frequently associated with successful treatment at each hospital. Figure 1 shows a possible result from such an algorithm.

Note that, in this illustration, the only variables that affect the difference between the two hospitals are gender, age, and diabetes (in males only). Many other variables, including obesity, other medical conditions, and other age categories, might affect the outcomes for individual patients, but if they do not affect the *difference* between the two hospitals' overall outcomes they are not part of the tree. The conclusion we can draw from Figure 1 is that Hospital A is superior for older women, while Hospital B is preferable for

younger men with diabetes. For all other categories of patients, there is no significant difference between the two hospitals. Because the tree differentiates between patient types, we call its information *patient-specific information*. Our hypothetical firm could use this information to designate each hospital as a COE for the patients for whom it performs better, while allowing other patients to choose whichever hospital is closer (and cheaper). By using patient-specific information to designate COEs, the firm will achieve better clinical outcomes and lower travel costs than it would by sending all patients to a single COE determined by population-average information.

A Case Study of Cardiac Procedures

Of course the above example is just an illustration. But actual data really do indicate this kind of patient-specific response to treatment at different hospitals. We used public data from the State of

New York on the outcomes of the thirty-five New York hospitals that performed open heart surgery between 2008 and 2012 to analyze their results with six cardiac procedures. These procedures, and the comorbidities we considered, are listed in Table 1. To carry out our analysis, we first generated trees for each of the 535 pairwise comparisons between these hospitals and then (for display purposes) translated these trees into assessments of the average complication score of each hospital as statistically better, worse, or the same as the state average.

Figure 2 is a graphic summary of our results, showing that the top

hospital achieves superior results for all patient types, the bottom hospital is worse than average for almost all patients, and the intermediate hospitals perform variably for different types of patients. Interestingly, the performance of some hospitals is very uneven. Hospital 11, for example, has average performance for most patients but significantly above average performance for all patients with hypertension. This achievement implies that a nearby community hospital can treat some patients as well as, or perhaps better than, a distant elite hospital.

A nearby community hospital can treat some patients as well as, or perhaps better than, a distant elite hospital.

The Power of Patient-Specific Information

To illustrate how a firm can use the patient-specific information summarized in Figure 2 to improve a COE program, we ran a simulation using the 2008-2012 data of New York cardiovascular patients. We assumed that patients select a hospital based on the quality of the hospital, as measured by quality adjusted life years (QALYs), which consider both the length and the quality of patients' lives after their procedures, as well as the distance to the hospital (which correlates with travel cost), and any financial incentives (e.g., travel subsidy, reduced co-pay, etc.) offered by their firm to offset the cost and inconvenience of traveling to a COE.⁸

In the first run of our simulation we assumed that the firm uses population-average information to identify COEs and patients use the same data to evaluate the risk of complication at each hospital. That is, the firm computes each hospital's O/E ratio for outcome measured in QALYs and designates the hospital with the highest average QALY as the

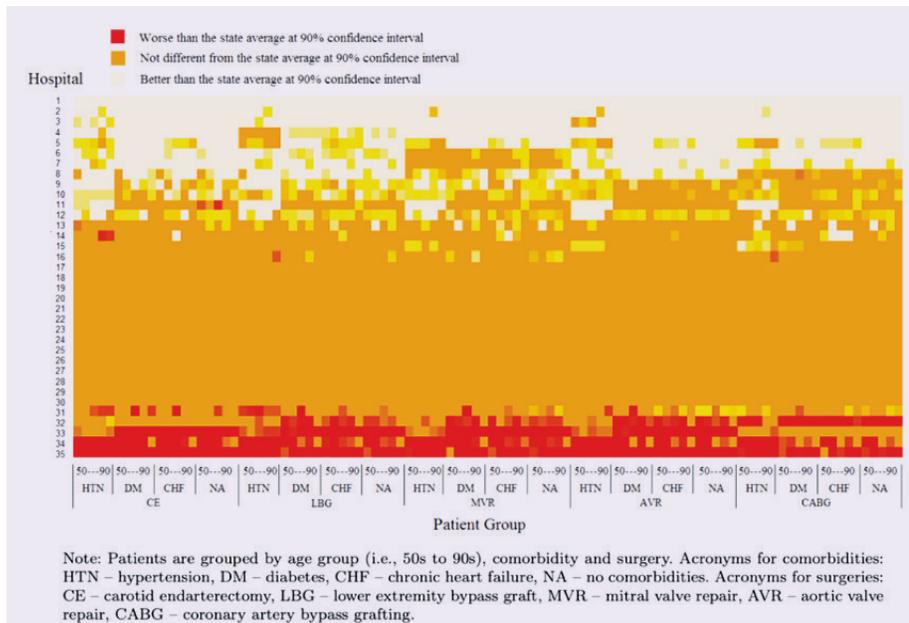


FIGURE 2: Comparison of New York hospitals for cardiovascular procedures, 2008-2012.⁷

Abbreviation	Cardiovascular Procedure	Abbreviation	Comorbidity
AAA	Abdominal Aortic Aneurysm	HTN	Hypertension
CE	Carotid Endarterectomy	DM	Diabetes Mellitus
LEG	Lower Extremity Bypass Graft	CHF	Congestive Heart Failure
MVR	Mitral Valve Repair/Replacement	NA	No Comorbidities
AVR	Aortic Valve Repair/Replacement		
CABG	Coronary Artery Bypass Graft		

TABLE 1: Cardiovascular surgeries performed in New York hospitals

COE for all patients. It also provides patients with the O/E ratios for all hospitals, so they can compare the expected QALY at their local hospital to that at the COE. With no incentives, patients for whom the QALY difference is large are likely to choose the COE, while patients for whom it is small will not. The more generous the incentives, the more patients will choose the COE, which will usually lead to better patient outcomes, but at a higher cost to the firm.

However, as we have noted, the population-average information from O/E ratios does not accurately represent the relative performance of hospitals for individual patients. In the second run of our simulation we therefore assumed that the firm and patients use patient-specific information generated by the described tree-based machine learning approach. By producing trees like the one shown in Figure 1, we provided both the firm and the patients with outcome comparisons of each hospital with every other hospital. We then designated the best overall hospital for each patient as the COE for that patient. Because such comparisons depend on patient characteristics, however, a patient's COE and the difference between expected QALY at the COE and the patient's local hospital naturally differ from patient to patient. Nevertheless, by using their own specifically tailored information, each patient could decide whether or not the health benefits of their COE outweigh the travel disbenefit (with any incentives factored in). We found that different patients will choose to travel when patient-specific information is used in place of population-average information, and that they will travel to different COEs. Because patients gain a more accurate understanding of their individual risks from patient-specific information, those who travel to

COEs will, on average, gain more clinical benefit from doing so.

The bottom line is that, by using machine learning to transform observational data into patient-specific information about hospital performance, firms can both save money and improve the health of employees in their COE programs.

Figure 3 illustrates the tradeoff between health outcome and travel incentive under population-average and patient-specific information. In both cases, by spending more money on incentives to encourage patients to be treated at a COE, a firm can improve their health outcomes by reducing the risk of complication, mortality, and readmission. However, because patient-specific information allows the firm to more accurately define COEs and guide the right patients to them, its use will lead to better health outcomes at any level of incentive investment. A firm can therefore improve health

outcomes while holding incentive spending constant (option A), reduce incentive spending such that health outcomes using patient-specific information end up the same as when population-average information is used (option B), or reduce incentive spending by a smaller amount while still improving health outcomes (option C). The bottom line is that, by using machine learning to transform observational data into patient-specific information about hospital performance, firms can both save money and improve the health of employees in their COE programs.

The Bottom Line of COE Programs

It is not surprising that more precise patient-specific information helps patients to make better health decisions and consequently allows firms to spend more efficiently on incentives. But will using this information also reduce the firm's health care costs? Because it depends on factors ranging from a given procedure's detailed clinical outcomes to the nature of the contracts between the firm and its health care providers, this is a complex question. What is clear is the ways in which such information can be used to increase cost efficiency.

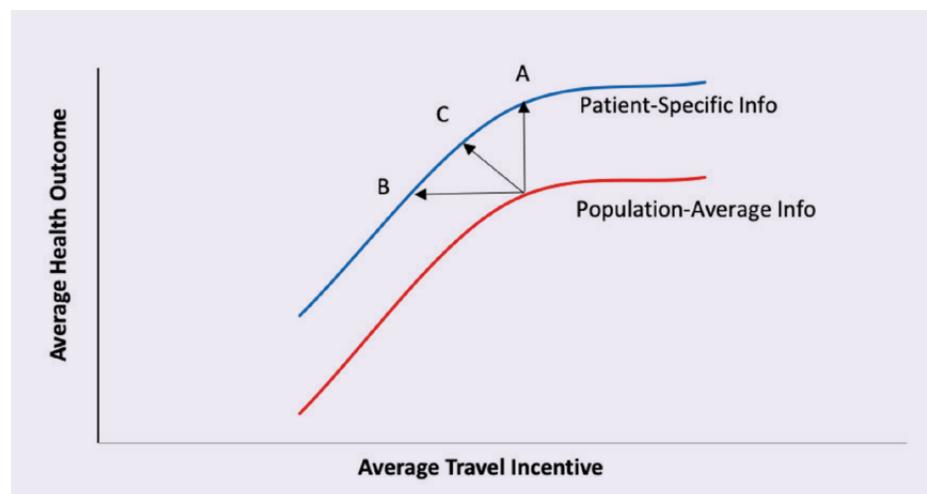


FIGURE 3: The impact of patient-specific information on average clinical outcome and travel incentive.

First, there are situations in which, even with a special negotiated price, the cost of a given procedure at a COE is higher than at a non-COE. From 2015 to 2018, for example, Walmart paid 8 percent more per patient for spinal surgery at a COE than at a non-COE hospital.⁹ If the outcome data for spine surgery exhibit a pattern similar to that shown for cardiovascular surgery in Figure 2, Walmart is probably paying this additional cost, in addition to travel incentives, for some patients whose health does not benefit. Using patient-specific data would enable Walmart to reduce costs without hurting patient outcomes.

Second, although Figure 2 shows that Hospital 1 is better than average for all patients, it is still not necessarily the best choice for all patients. Indeed, a finer analysis than we can display in a simple heat map reveals that Hospital 1 is the preferred choice for only some patients. If the same is true for spinal surgery patients then even though patients are, on average, better off at the COE, some are worse off. And even when the cost of the initial procedure remains the same when the firm does not send such patients to the COE, it may save on future health care costs by avoiding complications or recurrences.

By using the patient-specific analysis we have described here, firms could determine which patients will benefit from being diagnosed at the COE, as well as which will benefit from treatment at the COE.

Finally, superior diagnostics are a key factor that can offset the higher

cost of surgery at the COE. They ensure that fewer patients undergo surgery at all, allowing some to instead receive effective but less invasive (and costly) treatment such as oral drugs, injections, or physical therapies. It is not necessary, however, to bundle diagnosis and treatment. By using the patient-specific analysis we have described here, firms could determine which patients will benefit from being diagnosed at the COE, as well as which will benefit from treatment at the COE. There may be groups for whom diagnosis at the COE is helpful but treatment at the COE is not. In such cases, the firm is spared the cost of unnecessary or inappropriate surgeries by having patients diagnosed at the COE and spared the extra costs associated with remote treatment by having patients treated at local hospitals.

Toward More Effective and Efficient Health Care

Leveraging observational data and using machine learning algorithms to generate patient-specific information is straightforward in theory. In practice, Mies van der Rohe's belief that, "God is in the details" applies. Observational data are messy and fragmented. Machine learning algorithms are sensitive to tuning parameters. Given these complexities, does it make sense for Walmart, Lowe's, GE, Boeing, PepsiCo, and every other firm with a COE program to independently analyze hospital performance? They are already in the health care business; do they need to take on the big data analytics business as well?

We suspect not. Since firms with COE programs are not in competition over health care, they stand only to gain from sharing data and analytic results. One means of doing so would be for a third party to compile and analyze the data needed to generate patient-specific

hospital performance statistics. A public facing organization such as the federal Department of Health and Human Services' Centers for Medicare & Medicaid Services (CMS) or the private non-profit Leapfrog Group are already compiling the hospital and patient data needed for these analyses. Both use the data to populate websites that compare hospital performance.¹⁰ Their lists of procedures and sets of outcome metrics are ever expanding. Unfortunately, to date, both sites offer only population-average information.

Since firms with COE programs are not in competition over health care, they stand only to gain from sharing data and analytic results.

However, by applying machine learning and data analytics to the same data, these organizations could readily generate patient-specific statistics. If they did, their hospital performance summaries would no longer be simple lists that look the same to everyone. Instead, their websites would generate customized hospital statistics informed by patient profiles composed of information about sex, age, existing medical conditions, prior procedures, and other characteristics that might influence a patient's choice of hospital. By doing so, they would better serve their patients.

Meanwhile, firms could use those sites to guide their COE programs by processing the same patient-specific data against a list of candidate hospitals and using the results to generate a report that identified COEs and which patients would be best treated in them. Another option, if public facing orga-

nizations do not transform their sites to offer patient-specific data, is for firms with COE programs to form a consortium or to collectively engage a third-party administrator to generate and share the necessary data. Any collaborative approach will reduce the cost of collecting, organizing, and analyzing the data which will make COE programs more efficient and effective.

COE programs are a promising innovation which allow large firms to both improve the quality and moderate the cost of employee health care. They can achieve both of these goals more effectively by using machine learning and analytics tools to mine the vast store of observational health care data. And in so doing, they will catalyze the big data revolution which will reduce the cost and improve the quality of health care for all of us. ■

Author Bios



Wallace J. Hopp is the C.K. Prahalad Distinguished University Professor of Business and Engineering at the Ross School of Business, University of Michigan. His research focuses on manufacturing, supply chain, and healthcare systems. whopp@umich.edu



Jun Li is an Associate Professor at the Ross School of Business, University of Michigan. She conducts research in data analytics, health-care management, revenue management, and supply chain management. junwli@umich.edu



Soroush Saghafian is an Associate Professor at the Harvard Kennedy School of Government, Harvard University. His current research emphasizes the development and application of stochastic system models to problems in health-care and operations management. Soroush_Saghafian@hks.harvard.edu



Guihua Wang is an Assistant Professor at the Naveen Jindal School of Management, University of Texas at Dallas. His current research revolves around the application of empirical econometrics and machine learning to healthcare operations issues. guihua.wang@utdallas.edu

Endnotes

1. KFF 2019. 2019 Employer Health Benefits Survey. Kaiser Family Foundation. <https://www.kff.org/health-costs/report/2019-employer-health-benefits-survey/> [accessed 05/15/2020]
2. In a survey of large, self-insured employers, the Society for Human Resources Management (2018) found that the percentage of firms contracting directly with COEs increased from 12 percent to 18 percent from 2018 to 2019. SHRM. 2018. For 2019, Employers Adjust Health Benefits as Costs Near \$15,000 per Employee. Society for Human Resources Management. <https://www.shrm.org/resourcesandtools/hr-topics/benefits/pages/employers-adjust-health-benefits-for-2019.aspx> [accessed 5-15-20].
3. Brescia, A., Paulsen, M., Watt, Rosenbloom, L., Wisniewski, A., Li, J., Wang, G., Likosky, D., Hopp, W., Bolling, S. 2020. Economic analysis and long-term follow-up of distant referral for degenerative mitral valve repair. *The Annals of Thoracic Surgery* (forthcoming).
4. Wang, G., J. Li, W.J. Hopp, F. Fazzalari, S. Bolling. 2015. Cost-effectiveness of referring patients to centers of excellence for mitral valve surgery. Ross School of Business, University of Michigan, Ann Arbor, MI.
5. Angrist, J., Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
6. Athey, S., G. Imbens. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Science* 113(27): 7353–7360
7. Athey, S., J. Tibshirani, S. Wager. 2019. Generalized random forests. *Annals of Statistics* 47(2), 1148-1178. Wang, G., J. Li, W.J. Hopp. 2019. An instrumental variable tree approach for detecting heterogeneous treatment effects in observational studies, Ross School of Business, University of Michigan, Ann Arbor, MI.
8. Wang, G., J. Li, W. Hopp, F. Fazzalari, S. Bolling. 2019. Using patient-specific quality information to unlock hidden healthcare capabilities. *Manufacturing & Service Operations Management* 21(3):582-601.
9. Woods, L., J. Slotkin, M. Coleman. 2019. How employers are fixing health care. *Harvard Business Review*. March 13, 2019.
10. See: <https://www.medicare.gov/care-compare/?providerType=Hospital&redirect=true> and <https://ratings.leapfroggroup.org/>