YELP FOR HEALTH

USING THE WISDOM OF CROWDS TO FIND HIGH-QUALITY HOSPITALS

Paul Howard
Senior Fellow

Yevgeniy Feyman
Adjunct Fellow

MI

About the Authors

Paul Howard is a senior fellow at the Manhattan Institute, where he is director of health policy. He is a contributor to The Apothecary (the *Forbes* blog on health care policy and entitlement reform) and a regular columnist for The Morning Consult. He is coauthor, with Peter Huber, of *Unlocking Precision Medicine* (Encounter, 2016).

Howard has written on a wide variety of medical-policy issues, including FDA reform, biopharmaceutical innovation, consumer-driven health care, and Medicare and Medicaid reform. He is often quoted on health-care issues, and his work has appeared in such publications as Bloomberg View, *Wall Street Journal*, *National Affairs*, *USA Today*, RealClearPolitics, *New York Post*, *Investor's Business Daily*, *Health Affairs*, and FoxNews.com. He is a member of MI's Project FDA.

Howard was part of the health-care-policy advisory group for Mitt Romney's 2012 presidential campaign, has testified before Congress, and, from 2013 and 2016, served on an expert panel as a judge for Celgene's Innovation Impact Awards. He joined MI in 2000, as deputy director of its Center for Legal Policy, where he edited research papers, managed legal policy analyses, and organized conferences. He holds a B.A. from the College of the Holy Cross and a Ph.D. in political science from Fordham University.

Yevgeniy Feyman is an adjunct fellow at the Manhattan Institute and a senior research assistant in the department of health policy and management at the Harvard T. H. Chan School of Public Health. His prior research has focused on subjects including the cost of health care and entitlement reform, drug pricing reform, and hospital consolidation. Feyman's current work is focusing on the dynamics of the Medicare Advantage market. He was previously the deputy director of health policy at the Manhattan Institute.

In 2013, Feyman, with colleagues Avik Roy and Paul Howard, released the Obamacare Impact Map, a state-by-state look at the effects of the ACA. The map has been cited numerous times on Capitol Hill for its help in understanding the law's effect on Americans.

Feyman has written for publications including *National Affairs*, the *New York Times's* Room for Debate, *Boston Globe's* STAT, *Health Affairs*, and *Politico*. He has spoken on numerous radio and TV shows and is a contributor to The Apothecary. Feyman holds a B.A. in economics and political science from Hunter College of the City University of New York.

Acknowledgments

Contents

Executive Summary

Online tools like ZocDoc, HealthGrades, and Yelp have become popular among people who search for digital information about physicians and hospitals. Yelp, one of the most widely used platforms, allows patients to rate health-care providers through a five-star rating system that can include narrative text reviews. In 2015, Yelp partnered with ProPublica to publish average wait times, readmission rates, and the quality of communication scores for more than 25,000 hospitals, nursing homes, and dialysis clinics. (Yelp employs an algorithm to prevent fraudulent, duplicative, or provider-generated reviews from influencing rating scores.)

According to some research, Yelp reviews correlate with Medicare surveys such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). We don't know, however, how accurate these reviews are for identifying *quality* doctors and hospitals.

This paper examines whether there is a correlation between Yelp reviews of New York State hospitals and objective measures of hospital quality, including readmission to a hospital after a surgical procedure and mortality rates after hospital treatment for certain conditions (such as heart attacks) or procedures (such as stomach surgery). We find that higher Yelp ratings are correlated with better-quality hospitals and that they provide a useful, clear, and reliable tool for comparing the quality of different facilities as measured by potentially preventable readmission rates, a widely accepted metric.

We do not argue that Yelp alone is, or can be, the only guide to quality hospitals. However, when people can choose where they will obtain care—as do patients with traditional Medicare coverage for elective or planned surgeries, or when consumers can choose among insurance options—Yelp ratings can provide a helpful guide. Indeed, when patients seek out specialists for surgical or other hospital procedures, these specialists' hospital privileges could factor in their decisions, supplementing the more traditional system of referrals. This tool will get even better at helping patients identify quality institutions and physicians as the number of Yelp's hospital and provider reviews increases.

New York State policymakers, private and public employers, and Yelp itself thus have a tremendous opportunity to help consumers navigate the health-care system by considering the following recommendations:

- ✓ Help make Yelp scores more visible when consumers are making important decisions about health-care coverage—for instance, when choosing among competing insurers' hospital networks on New York State's health-insurance exchange.

- ✓ Link objective, simple quality metrics onto the Yelp review page for hospitals to allow patients with specific concerns to access more detailed information that would complement and better inform Yelp quality ratings.

- ✓ Fund targeted "hackathons" that find ways to make Yelp and other social media reviews more accessible to high-needs, vulnerable populations—including caregivers for the frail, elderly, non-English-speaking, or low-income minority populations.

By disseminating neutral, clear signals about basic hospital quality, social media tools can also improve the ability of higher-quality hospitals to compete to attract market share, leading to more lives saved and more costs avoided for patients, taxpayers, and employers.

YELP FOR HEALTH

USING THE WISDOM OF CROWDS TO FIND HIGH-QUALITY HOSPITALS

Introduction

Hundreds of thousands of New Yorkers in both employer-based and public-exchange health plans are being asked not only to make greater financial contributions to the cost of their care but to wisely choose when and where they access doctors and hospital-based care. A potentially promising, low-cost tool for helping patients make informed choices is social media platforms that host consumer-generated reviews, including ZocDoc, HealthGrades, and Yelp, one of the most popular of such platforms.

The key question is whether consumer-generated reviews are reliable indicators of physician or hospital quality. We have tested the correlation between Yelp reviews of New York State hospitals and objective measures of quality, including how often patients are readmitted to a hospital after a surgical procedure and mortality rates after hospital treatment for conditions (such as heart attacks) or procedures (such as stomach surgery).

Our findings suggest that higher Yelp ratings are associated with better-quality hospitals, using the metric of potentially preventable patient readmissions. (The relationship of Yelp ratings to mortality rates after hospital treatments is less clear, and we discuss possible reasons why this is so below.) The upshot is that Yelp ratings have the potential not only to empower patients with concise, reliable information about hospital performance but to help redirect market share to higher-quality hospitals, which would save lives and lower costs.

Consumer Trade-Offs: Choice, Cost, and Quality

Balancing choice and cost continues to be a conundrum for health-care plans. When they are surveyed, Americans consistently say that they want a wider array of insurance choices, including plans covering more benefits or larger networks of doctors and hospitals.[1] Research, however, has repeatedly identified wide variations in provider cost and quality, even after adjusting for the severity of a patient's illness.[2] Medical errors have been estimated to account for more than 200,000 American deaths annually.[3] And some estimates conclude that more than 30% of U.S. health-care spending may be on care that does not improve health outcomes—representing hundreds of billions of dollars in wasted money.[4]

Insurers have responded by engaging consumers to shop for value when they need care—driving them toward lower-cost or higher-quality providers, including narrow networks and tiered networks that offer lower co-pays for patients who stay in network—and by providing some measures of quality for physicians and/or hospitals in their networks.

The rise of high-deductible health plans, for instance—covering nearly one in three Americans in 2016[5]—means that consumers pay for an increasing amount of care out of pocket, providing another reason to seek out low-cost, high-quality options. America's Health Insurance Plans, an industry trade group, estimates that more than 600,000 New Yorkers were enrolled in high-deductible health plans (HDHPs) in 2015.[6]

FINDING RELIABLE, EASY-TO-UNDERSTAND INFORMATION ABOUT QUALITY IS A **SIGNIFICANT CHALLENGE FOR MOST PEOPLE**

Beyond deductibles, consumers who want more choice will often have to pay more for it. On average, 63% of all plans available on the Affordable Care Act's (ACA) public exchanges are "gatekeeper"-style health maintenance organizations (HMOs) or exclusive provider organizations (EPOs), which limit coverage to preferred physician or hospital networks, with greater cost-sharing for out-of-network providers. In some cases, patients bear 100% of the costs if they go out of network.[7]

Employers are also experimenting with network designs that funnel employees toward providers who deliver high-quality services—determining quality by metrics such as fewer hospital readmissions after cardiac surgery or fewer ER visits for employees with chronic conditions like diabetes. Readmissions have been iden-

tified as an important signal of hospital quality. MedPAC (the Medical Payment Advisory Commission) estimates that 17%–20% of Medicare patients discharged from a hospital were readmitted within 30 days,[8] and research has found that as many as 76% of these readmissions could be prevented by better hospital coordination and planning.[9]

Nearly 40% of large employers (5,000+ employees) offer these tiered or "high-performance" plans, a figure that is expected to rise if the ACA's tax on high-cost health plans is not repealed.[10] High-performance networks usually come with little or no cost-sharing and include incentives for providers who meet performance benchmarks. The Employee Benefit Research Institute also reports a sharp growth in New York in narrow networks, especially among small employers.[11]

In short, consumers continue to have choices among competing providers but are increasingly having to balance cost versus provider quality. But finding reliable, easy-to-understand information about provider quality is a significant challenge for most people.

Insurers and other third parties (including the federal and state governments) have developed tools that offer patients some measures of quality and cost. The problem is that they are not consumer-friendly: each of these tools requires patients to sort through a dizzying array of provider choices, including doctors, urgent-care centers, and hospitals, each with potentially different health risks, benefits, and cost implications for patients and payers. In addition, quality metrics, such as 30-day readmission rates or various process measures, only add to the complexity and may generate confusion for patients.[12]

If patients cannot easily find trusted and accurate information about provider quality, they are likely to become overwhelmed or confused by the options and will continue to seek out providers based on anecdotes or brand recognition because they can't separate low- from high-quality options. New Yorkers will continue to pay for low-quality care that leads to excess spending, poor health outcomes, and perhaps even needless deaths. High-quality providers without a well-known brand name will struggle to gain market share, compared with lower-quality or more expensive options.

However, if social media platforms are reliable (if they reflect objective information about hospital and pro-

vider quality), policymakers can incorporate them into sites where consumers have to make important decisions about care—for instance, when purchasing coverage through New York's health-insurance exchange. Policymakers could include Yelp reviews of insurers' in-network hospitals, providing an important and easily understood signal of network quality.

Even patients in fully subsidized plans like Medicaid could benefit from these ubiquitous tools. While they get care at little or no cost, they still face decisions about where to seek the care that will give them or their loved ones the best chance of a good outcome, the shortest wait in an emergency room, culturally appropriate care, and the smallest chance of a negative outcome such as a hospital-acquired infection, readmission, or complications from surgery.

Using Technology to Level the Playing Field Between Patients and Providers

Skepticism about consumerism in health care dates back at least to Kenneth Arrow's seminal 1963 essay, "Uncertainty and the Welfare Economics of Medical Care."[13] Arrow identified five ways in which health care seemed to diverge fundamentally from other markets, but perhaps his most significant critique was of the "information asymmetries" that he articulated between the consumers of health care (patients) and the providers of health care (physicians).

Arrow noted that doctors, thanks to their highly specialized, expensive, and time-consuming training, simply knew much more about the care required by a patient with a given condition, including the relative effectiveness of various treatment options. Given these inherent advantages, patients were bound to defer to the judgments of their physicians. Complicating matters still further, providers in a fee-for-service environment (as was the case when Arrow was writing, and as is still largely the case today) have a financial incentive to recommend more expensive or more intensive services that may be more profitable. The checks on price that consumer choice and producer competition generate in other markets are missing, or largely missing, in health care.

Today, the challenges that Arrow identified remain, but the availability of new, easily accessible web-

based technologies allows essentially cost-free access to rapidly updated information on health-care conditions, treatment options, and provider quality.

The federal government, for instance, is making a concerted effort to make more provider performance available to consumers through its "Star Ratings" (easily accessible online) for various facilities, including hospitals, nursing homes, and dialysis centers. In 2015, Congress passed legislation to reform Medicare physician payments and create new value-based reimbursement tools that tie payments explicitly to quality measures or to participation in a non-fee-for-service program (like an Accountable Care Organization).

While many of these transparency tools are not yet widely used and contain information that is not easily digestible to consumers, they do represent a sea change from the conditions that existed when Arrow wrote and a resource for developing more objective quality ratings.

Health economists have also developed a more nuanced understanding of the role of competition and patient choice in health-care markets, especially as it relates to hospital consolidation and pricing. There is a consensus in the economics literature that hospital competition in less consolidated markets not only helps keep price increases in check but can raise the quality of care delivered to patients for at least some selected conditions.

Research has found that, in Britain's National Health Service, giving patients a choice among competing publicly financed hospitals led to improved patient outcomes for heart attacks.[14] Two large literature reviews[15] found that, in general, patients in less concentrated hospital markets had lower mortality rates for a variety of conditions.

Competition in other markets helps shift market share over time from low-quality (and/or less efficient) producers to higher-quality ones. This means that the quality for a given price increases across the market, and less efficient providers exit the market entirely. For instance, Toyota's ability to deliver exceptionally high-quality cars at competitive prices forced domestic U.S. car makers to retool their manufacturing plants to compete on both price and quality. Indeed, the inflation-adjusted price of a new car has barely budged since the mid-1990s.[16] But today's vehicles come standard with features—power windows, smartphone connectivity, touch-screen displays—that simply didn't exist or were available only in luxury vehicles in the 1990s.

If information asymmetries truly dominated the hospital market, we *shouldn't* see higher-quality hospitals gain market share over time. But a 2016 study by MIT and Harvard researchers of Medicare data for heart attacks, congestive heart failure, pneumonia, and knee and hip replacements found that competition played a large role in shifting hospital market share and even improving health outcomes.[17] Using clinical outcomes, process measures, and patient surveys, this study (published in the *American Economic Review*) found "robust evidence that higher performing hospitals—as defined either by the health outcome-based measures or the process of care measures—tend to have greater market share (i.e., more Medicare patients) at a point in time, and experience more growth in market share over time."

The correlation between hospitals and quality was highest for transfers from one hospital to another, compared with admissions from the emergency room. In short, according to this study, when patients and physicians had a choice about where to seek treatment, they transferred to a higher-quality hospital, suggesting an important role for consumer demand ("either by patients or by their surrogates") in driving market share to higher-quality hospitals. In fact, the trend toward higher-quality hospitals was responsible for as much as 25% of the decline in heart-attack mortality from 1996 to 2008. Heart failure and pneumonia also saw declines, although not as large.

The findings of this study are an encouraging signal that health-care consumerism can help drive patients toward better-quality providers even in traditional fee-for-service Medicare, where patients have virtually unlimited choice of hospitals and little financial skin in the game (thanks to Medigap supplemental insurance plans that offer broad coverage of cost-sharing in the Medicare program). Still, the information on hospital or provider quality collected by government agencies may have serious drawbacks, from a consumer perspective. It may be difficult to find or access, be excessively technical and be collected only after significant time lags, or not be directly relevant to the particular needs or interests of the patient.

Social Media and the Wisdom of Crowds in Health Care

Social media and the wisdom of the crowds that it implies—the aggregation of growing numbers of patient experiences in real time—may be able to reduce rating confusion and increase the use of rating scores without sacrificing significant reliability. In fact, as more consumers use social media platforms and comment on them, they create a real-time feedback loop on consumers' health-care experiences with physicians and hospitals that is likely to improve over time. Already, sites like Yelp are becoming the go-to tools that consumers use to choose one doctor or facility over another.

Surveys suggest that 75% of Americans with internet access searched for medical information online in 2012. These people are increasingly turning to social media, as well. Some 16% have had online interactions with other patients presenting with the same health conditions, and about 25% have read or watched someone else's experience about health issues.[18]

What we don't know is how accurate these reviews are in terms of identifying *quality* doctors and hospitals. Is it possible to obtain reliable information on social media about whether a physician or hospital is high-quality or not? Even well-structured surveys of patient experience—like the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)—have an uneasy relationship with outcomes. Prior research has found some relationship in this survey between patient satisfaction and objectively measured health outcomes—namely, lower mortality—but not with others.[19] Indeed, scores in one domain of the survey have been associated with an increased risk of readmission.[20] These somewhat paradoxical results suggest that evaluating widely used social media tools—which rely on a single composite rating generated from the experiences of many different consumers—is a priority. This evaluation is our next step.

Reviews on social media typically track one encounter and lack a full provider and patient context. Still, they arguably compare favorably to the pre-Internet status quo, where consumers relied on word of mouth from a small network of friends or family members to choose a doctor or hospital. Online health-care tools to evaluate doctors and hospitals have proliferated in recent years, but many, like the Leapfrog Group's Hospital Safety Grade and Medicare's Hospital Compare, don't appear to be highly trafficked, compared with social media options. Castlight Health's cost- and quality-transparency products are more comprehensive but are tailored to the needs of specific employers and are available only to those who have insurance through an employer.

Insurers rate providers individually, but the methodologies behind these ratings vary across insurance companies, are not easily understood by consumers, and

often lack subjective information that is important to consumers, such as assessments of physicians' bedside manner or the comfort of facilities.

Other tools, including the Health Care Cost Institute's Guroo, FAIR Health's NY Host, and Consumers Union's cost estimator, promise to significantly broaden consumer access to vital information about the price and quality of health services in New York but face similar challenges in terms of accessibility. Clearly, widely trafficked social media tools like Yelp can complement these objective quality efforts in terms of access and usability.

Yelp for Health

We focus on Yelp because of its size. Yelp has more than 140 million unique monthly visitors. It is famous for reviews of local businesses and restaurants, but Yelp also allows patients to rate hospitals, nursing homes, and dialysis clinics, based on its five-star rating system, along with a written narrative that reflects the user experience. Yelp teamed up with ProPublica in 2015 to publish government-collected data on average wait time, readmission rates, and quality of communication from 4,600 hospitals, 15,000 nursing homes, and 6,300 dialysis clinics.[21] It includes an algorithm to prevent fraudulent, duplicative, or provider-generated reviews from influencing rating scores.

To date, only a few studies have attempted to compare Yelp's provider reviews with other rating systems. In 2013, a study in *BMJ* (British Medical Journal) *Quality and Safety* attempted to compare Yelp hospital ratings with hospitals' overall score on Medicare's HCAHPS survey, individual HCAHPS scores for measures like pain control, and a limited data set on objective outcomes, including hospital mortality and 30-day readmission rates. It found that Yelp and HCAHPS scores were correlated, with higher Yelp scores indicating higher HCAHPS scores. The percentage of high ratings on Yelp and HCAHPS was significantly correlated with lower mortality for heart attacks, pneumonia, and hospital readmissions. The study concluded that "rater experiences for Yelp and HCAHPS may be similar, and that consumers posting ratings on Yelp may observe aspects of care related to important patient outcomes."[22]

An April 2016 study in *Health Affairs* also compared Yelp reviews with HCAHPS. It found that overall hospital ratings were strongly correlated across HCAHPS and Yelp (conditional on at least three Yelp hospital ratings, compared with five in the earlier study). By ana-

lyzing the narrative reviews included with Yelp ratings, this study also found that Yelp covered far more areas of interest to patients than HCAHPS while covering most (seven out of 11) of HCAHPS domains. These included the cost of hospital visits, insurance and billing, additional testing, facilities, nursing quality, care of family members, staff quality, quality of technical care, and specific types of medical care. HCAHPS suffered from significant lag times between surveys and low rates of utilization. Yelp reviews were written "in real time and often written by patients for patients" and "perceptions of what matters most to them can change over time."[23] Yelp reviews also capture the opinions of caregivers, which may be particularly salient in cases where patients are frail, elderly, or unable to complete a survey or adequately advocate for themselves.

Writing in the *New York Times*, Aaron E. Carroll, a physician and health-policy researcher, concluded that research on Yelp and hospital choice showed that "patients reward quality on their own—when they can. We might just need to make it easier for more of them to do so."[24]

While both studies addressed the consistency of patients' perception of quality across Yelp and HCAHPS, and even identified additional domains of interest, they were limited. The 2013 study, for example, focused on objective outcomes but included only one year of outcome data. Moreover, the outcome data were limited to Medicare patients, who may be less likely to review hospitals on a social media platform.

Methodology and Data

Our comparison of Yelp ratings with objective measures of hospital quality used several sources to construct an analytic sample and limited the analysis to New York State. Hospital-level outcomes and some of our control variables came from New York State's Statewide Planning and Research Cooperative System (SPARCS) database. (Other controls came from the American Hospital Association Annual Survey.) SPARCS's administrative database contains all inpatient discharges from all facilities, including hospitals and nursing homes, in New York State. Our sample encompassed acute-care hospitals in New York State from 2009 through 2014. Children's and women's hospitals, rehabs, and other facilities were excluded.

Our outcome variables, sourced from the SPARCS database, included three measures of hospital quality

risk-adjusted for differences in various patient characteristics: potentially preventable readmission (PPR) rates, mortality for selected conditions, and mortality for selected procedures. The latter two of these measures have been endorsed by the respected National Quality Forum and represent a set of useful, relevant measures of hospital quality, and the first is widely used as another barometer of hospital quality.[25] All three measures were obtained from data sets created by the New York State Department of Health (NYSDOH) from the SPARCS database. (The Appendix to this report has a more detailed discussion of these measures.)

Measures of readmission are particularly useful because experts have widely agreed that many readmissions are preventable, costly, and associated with worse outcomes for patients. The PPR measure attempts to identify which readmissions are preventable (versus those that aren't, or are necessary) and is calculated by 3M Health Information Systems for the NYSDOH.

We also used SPARCS data to identify several hospital-level characteristics that are likely to be associated with the quality measures that we are examining. The measures we use are risk-adjusted at the patient level—that is, they are adjusted for various patient characteristics that may influence clinical outcomes. We also add controls for the racial makeup of each hospital's patients, the hospital's share of Medicaid patients, and the hospital's share of surgical patients. A larger share of low-income Medicaid patients, who may have greater health challenges than more affluent populations, may make a hospital's performance look worse than it is in reality. (Medicare's Hospital Readmission Reduction Program will soon begin factoring into account socioeconomic status.) Similarly, a hospital with surgical admissions making up a large share of total discharges may mean higher patient complexity, which would result in worse (unadjusted) outcomes on average. These additional controls allow us to account for the effects of socioeconomic status, race, and patient severity at the hospital level, all of which can influence patients' clinical outcomes.

We obtained additional hospital-level characteristics for a subset of our data from the American Hospital Association (AHA) 2014 Survey. From AHA, we collected data on whether the hospital is part of a health system, its size (by the number of beds), and its resident-to-bed ratio.

Lastly, we obtained yearly average reviews of New York State hospitals from Yelp (over the same sample period). Our sample was restricted to "trusted ratings," which have been vetted by Yelp's algorithm for identifying duplicate or possibly fraudulent reviews.

We were not able to obtain Yelp ratings or AHA data for the full universe of New York State hospitals. Our sample of SPARCS facilities varied slightly each year—137 in 2009 and 2010, 139 in 2011, and 138 for the remaining years. Our final sample across all years included Yelp ratings for 52.7% of these facilities (436 facility-year observations in total). Incorporating AHA data brought our sample down to 45.2% of the SPARCS data (374 facility-year observations). For additional information on methodology and the graphic representation of the study's results, see the **Appendix**.

Results and Discussion

The results of our analysis suggest that Yelp ratings are useful and reliable, on their own, for comparing the quality, *as measured by risk-adjusted, potentially preventable readmissions* (PPR), of different hospitals. In both contemporaneous (same year) and lagged models (prior year's rating), the correlation as measured by the R-squared is close to 0.2 for this measure between hospitals—which suggests that within a given year, the Yelp rating helps explain roughly one-fifth of the variation in PPR between hospitals. Additionally, the Yelp rating has statistical significance, and its coefficient is directionally what we might expect.

YELP RATINGS ARE USEFUL AND RELIABLE FOR COMPARING THE QUALITY, AS MEASURED BY RISK-ADJUSTED, POTENTIALLY PREVENTABLE READMISSIONS (PPR), OF DIFFERENT HOSPITALS

Simply put, a lower PPR—which signifies a higher-quality hospital—is associated with a higher Yelp rating. Thus, if patients were to use the Yelp rating to try to identify higher-quality hospitals, it's likely that they would be sent in the right direction and would probably benefit from higher-quality care.

PPR is not the only measure of hospital quality in our study, or elsewhere. The predictive power of the Yelp rating for mortality (mortality for a set of procedures, or patients' conditions) is low—that is, the Yelp rating explains relatively little variation in hospital outcomes. Paradoxically, hospitals in a given year with slightly higher Yelp ratings appear to have slightly higher mortality rates for procedures. It isn't clear why this is the case, though the smaller variation in mortality for con-

ditions and procedures, compared with PPR, may be one explanation.

When we controlled for the share of Medicaid patients and surgeries at the hospital level, however, the Yelp rating loses statistical significance even for PPR, while the overall explanatory power of the model increased. The controls that we selected may capture more variation in PPR than the Yelp rating does and also tend to contain similar "information" about a hospital's outcomes.

This finding about Yelp ratings compared with our controls does not mean that the Yelp rating isn't a useful measure for consumers looking to assess hospital quality. Indeed, Yelp ratings do have explanatory power on their own. But patients or researchers who want to explore hospital performance more closely will find that the hospital characteristics that we control for (in addition to the patient-level adjustments) are better tools for explaining hospital performance on our chosen outcomes. The limitation is that the typical patient will not be able to parse (or even access) the characteristics that we've controlled for. Yelp ratings, on the other hand, offer a single, easily understood measure that correlated well with hospital quality—and also contain other information of interest to prospective patients not captured by other surveys—and thus represent a very useful search tool for patients.

Yelp scores for health care are relatively new. The low sample size of Yelp ratings in earlier years (closer to the tool's launch) may also contribute to the relative weakness of Yelp ratings as a predictor of quality, compared with our chosen controls. The number of hospitals with Yelp ratings and the number of Yelp ratings at each hospital are likely to increase in the coming years, and social media tools will likely continue to evolve among patients as well as providers. These developments suggest that future research would be useful to understand whether Yelp ratings become even more useful predictors of hospital quality over time.

Barriers to Patient Choice

A patient's ability to choose a hospital may be constrained in ways that blunt the ability of social media tools to direct him toward the best institutions. Insurers are increasingly tightening their hospital networks (for instance, in HMOs or through tiered networks), in order to negotiate lower reimbursement rates in return for greater volume at preferred institutions. Hospitals

and hospital systems, in turn, are increasingly purchasing physician groups, to capture additional referrals and gain bargaining power against insurers. After hospitals acquire these groups, physicians may have admitting or referral privileges only at a single hospital (or hospital system), limiting patient choice.

These trends may narrow, but not eliminate, the importance of competition and choice among hospital facilities. Or they may widen the importance of competition. Patients will still have an opportunity to choose among insurer networks, particularly on public or private exchanges, and among specialists with different admitting privileges. An increased focus on the high costs and poor performance attributed to hospital consolidation may give more affordable community hospitals an edge in pricing in insurer negotiations—provided they can effectively document their quality. New tools like reference pricing, where an insurer sets a single price for a procedure for hospitals that agree to accept bundled reimbursement, can provide a baseline for competition among hospitals, potentially opening networks rather than narrowing them. And as patient satisfaction is increasingly tied to reimbursement measures through Medicare, the importance of social media tools as feedback loops will become more critical to hospitals as well.

Conclusion

Health-care providers have expressed serious reservations about hospital quality ratings as inconsistent across the platforms that provide the ratings, and as not properly accounting for the sickness or frailty of patient populations. Providers also express concerns that customer satisfaction surveys don't capture dimensions of care that reflect clinical or objective quality metrics, especially for hospitals that serve vulnerable populations.[26]

Based on our research into New York State hospitals and Yelp scores, along with earlier studies that explore correlations with HCAHPS and Yelp, we believe that those concerns are largely unfounded. Yelp scores are, in fact, good composite measures of hospital quality. Other metrics may explain *more* of the variation in hospital quality, but they are metrics that most patients are unlikely to seek out or to be able to understand. Given that Yelp ratings correlate with *risk-adjusted* measures, our research suggests that the baseline correlation is not simply an effect of healthier patients selecting into hospitals, or healthier patients writing reviews on Yelp.

The chief weakness of Yelp reviews may be that they are relatively new. As the number of reviews in New York State increases, the scores may become stronger predictors of hospital quality. Accordingly, here are three recommendations to consider:

1. Help make Yelp scores more visible. Insurers on New York State's health-insurance exchange are required to list up-to-date provider directories. Yelp hospital ratings can be included with these directories, helping consumers determine not only if their preferred hospitals are in-network but if they are of high quality. The ratings could also help reassure consumers that plans with narrower networks (or networks of community hospitals) are of equal quality as—or even of higher quality than—more expensive plans with larger networks.

2. Regulators should discuss with Yelp (and possibly other social media platforms) the ability to link simple quality metrics from SPARCS or other data sets for common procedures and conditions onto the Yelp review page for hospitals. Linkage would let patients get more detailed information that would complement and better inform Yelp quality ratings. While Yelp would retain a clear and concise star rating, consumers could easily access other data sets from the facilities' main Yelp page. The California Healthcare Foundation and Yelp are currently exploring a similar effort.[27]

3. Consider funding targeted "hackathons" that find ways to make Yelp and other social media reviews more accessible to high-needs, vulnerable populations—including caregivers for the frail, elderly, non-English-speaking, and low-income minority populations. These hackathons could become part of New York's E-health Collaborative. The collaborative already contains a digital accelerator, which helps health-tech companies hone their business models and connect with potential customers. Apps, translation software, or even the ability for peer-to-peer communications with patient navigators—trusted Yelp reviewers who are culturally competent—are all ideas that could be explored through an accelerator and then launched through for-profit or non-profit services.

In the long run, Yelp's health-care ratings are apt to become an even more valuable research tool than they are at present. They can assist other efforts to develop more innovative payment models that better align payment with quality and drive consumers to more efficient providers. Policymakers should remember that signals generated by services like Yelp remain most valuable when they are independent and user-generated. Regulation is apt to strip these tools of the agility that makes them valuable to patients—and to reformers who care about the triple aim of delivering better care at lower cost to more New York residents.

> IN THE LONG RUN, YELP'S HEALTH-CARE RATINGS ARE APT TO BECOME AN EVEN MORE **VALUABLE RESEARCH TOOL** THAN THEY ARE AT PRESENT

Appendix

The Study's Methodology

We ran a series of random-effect linear regression models to account for between- and within-hospital variation in outcomes and Yelp ratings. Errors were clustered at the facility level. Our models included year fixed effects as well as a set of controls to account for plausible, observable confounders. The primary predictor of interest was the contemporaneous and lagged average annual Yelp rating for each hospital. We examined the coefficient on and significance of Yelp ratings, as well as the R-squared from each model.

We further restricted the analysis to facilities with at least two Yelp reviews. This restriction reduced the sample size to 374 observations for contemporaneous models and 315 for lagged models. Adding AHA controls to the model reduced the sample further to 330 and 278 observations, respectively.

Outcome Measures

We used three separate measures of hospital-level outcomes derived from the SPARCS database. The first, potentially preventable readmissions (PPR), attempts to measure readmissions to the hospital within 30 days of discharge that might have been preventable.

This measure is calculated by first excluding some admissions, including those for metastatic malignancies, trauma, burns, some obstetric admissions and newborns, patients who left against medical advice, as well as transfers. The remaining admissions are considered "index" admissions, and "PPR chains" are calculated based on the number of PPRs following the initial admission. The number of PPR chains is then divided by the number of "at risk" admissions. To obtain a risk-adjusted measure, a separate expected PPR rate is calculated using a model that controls for various patient-level characteristics, including age, mental health status, and diagnosis. The risk-adjusted PPR is calculated by dividing the observed rate by the expected rate (higher is worse, lower is better) and multiplying by the statewide observed PPR rate.[28]

The other two measures of hospital-level outcomes were mortality for a select group of procedures and conditions. For procedures, these include eight measures of mortality for the following procedures: esophageal resection, pancreatic resection, abdominal aortic aneurysm repair, coronary artery bypass graft, craniotomy, hip replacement, percutaneous coronary intervention, and carotid endarterectomy.[29]

For conditions, these include six measures: acute myocardial infarction, heart failure, acute stroke, gastrointestinal hemorrhage, pneumonia, and hip fracture.[30] All measures are risk-adjusted for patient-level characteristics, and the final measure is a ratio of observed-to-expected rates, where expected rates are calculated using a similar statistical model to the one described above. All three measures were obtained directly from SPARCS data, calculated by the NYSDOH and 3M (for risk-adjusted PPR).

Mean Number of Yelp Ratings per Hospital Year
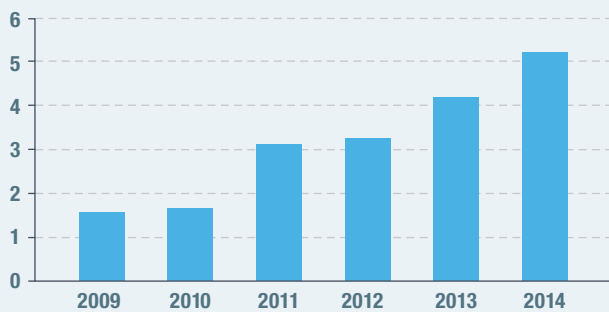
Source: Authors' calculations from Yelp data

Risk-Adjusted Potentially Preventable Readmissions (PPR)
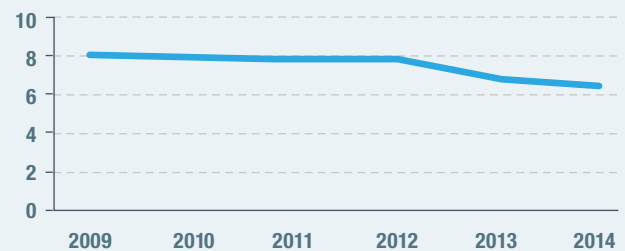
Source: Authors' tabulations from SPARCS data

Average Yelp Ratings
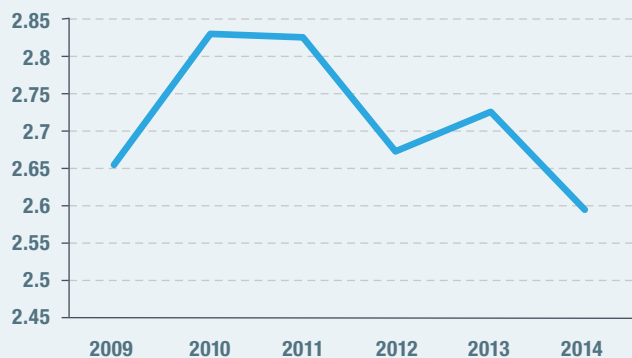
Source: Authors' tabulations from Yelp data

Summary of Control Variables

Summary of Control Variables	
Average Yelp Rating	2.70
Medicaid Share	23%
Black Share	17%
Surgical Share	22%
Resident–Bed Ratio	19%

Member of a Hospital System?	Average Yelp Rating
No	2.824257
Yes	2.685378

Bed Size	Average Yelp Rating
<=99	2.915727
100–499	2.729651
500+	2.71289

Source: Authors' calculations from Yelp, SPARCS, and AHA data

Lagged Model Results

Risk-Adjusted PPR	Model 1	Model 2	Model 3	Model 4
Lagged Rating	-0.109	-0.128***	-0.0548*	-0.0447
Year Controls		X	X	X
Medicaid Share			0.0842	-0.0566
Black Share			1.034**	1.022**
Surgical Share			-1.743**	-2.234**
Health System				-0.114
Resident–Bed Ratio				0.664***
100–499 Beds				-0.280
500+ Beds				-0.346
N	315	315	315	278
R2 Between	0.198	0.322	0.451	0.495
R2 Within	0.0141	0.447	0.462	0.480
R2 Overall	0.0279	0.366	0.464	0.492

Procedure Mortality	Model 1	Model 2	Model 3	Model 4
Lagged Rating	-0.00957**	-0.00667*	-0.00333	-0.00245
Year Controls		X	X	X
Medicaid Share			0.0914***	0.111***
Black Share			-0.00792	-0.0230
Surgical Share			-0.0778***	-0.0816**
Health System				-0.0122
Resident–Bed Ratio				-0.0429
100–499 Beds				-0.0111
500+ Beds				0.00281
N	315	315	315	278
R2 Between	0.0686	0.114	0.211	0.255
R2 Within	0.00999	0.147	0.144	0.135
R2 Overall	0.0552	0.108	0.205	0.254

Condition Mortality	Model 1	Model 2	Model 3	Model 4
Lagged Rating	-0.0223**	-0.0166	-0.0153	-0.0122
Year Controls		X	X	X
Medicaid Share			-0.153	-0.130
Black Share			-0.0209	-0.0550
Surgical Share			-0.319	-0.365
Health System				-0.147
Resident–Bed Ratio				-0.0622
100–499 Beds				-0.0133
500+ Beds				-0.0312
N	315	315	315	278
R2 Between	0.0152	0.00163	0.0875	0.110
R2 Within	0.0276	0.192	0.181	0.181
R2 Overall	0.00000538	0.0331	0.0853	0.121

* p<0.10 | ** p<0.05 | *** p<0.01

Contemporaneous Model Results

Risk-Adjusted PPR	Model 1	Model 2	Model 3	Model 4
Yelp Rating	-0.0604	-0.0709**	-0.0183	-0.0172
Year Controls		X	X	X
Medicaid Share			0.108	-0.0840
Black Share			1.180***	1.176***
Surgical Share			-1.780**	-2.297**
Health System				-0.0963
Resident–Bed Ratio				0.816***
100–499 Beds				-0.244
500+ Beds				-0.360
N	374	374	374	330
R2 between	0.194	0.240	0.381	0.455
R2 within	0.00944	0.474	0.484	0.497
R2 overall	0.0243	0.337	0.449	0.484

Procedure Mortality	Model 1	Model 2	Model 3	Model 4
Yelp Rating	0.00337	0.00484	0.00881***	0.00967***
Year Controls		X	X	X
Medicaid Share			0.101***	0.131***
Black Share			-0.00784	-0.0258
Surgical Share			-0.121***	-0.127***
Health System				-0.0205
Resident–Bed Ratio				-0.0489*
100–499 Beds				-0.0164
500+ Beds				-0.00645
N	374	374	374	330
R2 Between	0.0755	0.0201	0.167	0.250
R2 Within	0.0225	0.114	0.124	0.117
R2 Overall	0.0102	0.0292	0.158	0.235

Condition Mortality	Model 1	Model 2	Model 3	Model 4
Yelp Rating	-0.0174	-0.0163	-0.0160	-0.0183
Year Controls		X	X	X
Medicaid Share			-0.131	-0.0977
Black Share			-0.0314	-0.0668
Surgical Share			-0.308	-0.318
Health System				-0.145
Resident–Bed Ratio				-0.0527
100–499 Beds				-0.000649
500+ Beds				-0.0187
N	374	374	374	330
R2 Between	0.00277	0.0161	0.113	0.119
R2 Within	0.0111	0.150	0.141	0.149
R2 Overall	0.00606	0.0376	0.0949	0.124

Source: Authors' calculation

* p<0.10 | ** p<0.05 | *** p<0.01

Endnotes

1 Patients and consumers, when polled, consistently vote in favor of more choice among competing plans and providers. A Kaiser Family Foundation poll from February 2014 found that 51% of Americans preferred plans with broader provider networks, even if they had to pay more. Among the uninsured, however, most were willing to trade off leaner networks for lower premiums (Kaiser Health Tracking Poll, Feb. 2014). A July 2016 Rasmussen report noted that "in regular surveys since January 2013, nearly three out of four voters have said individuals should have the right to choose between different types of health insurance plans, including some that cost more and cover just about all medical procedures and some that cost less while covering only major medical procedures. Even more (85%) regularly say they should have the right to choose between plans with higher deductibles and lower premiums and others with lower deductibles and higher premiums." At the same time, when the costs of larger networks or pharmacy networks, as in the ACA exchanges or Medicare Part D, become fully visible in the form of higher premiums, plans with fewer choices are clearly popular choices among cost-sensitive consumers.

2 For a review of the literature, see Elliott Fisher et al., "Health Care Spending, Quality, and Outcomes: More Isn't Always Better," Dartmouth Institute for Health Policy and Clinical Practice, Feb. 27, 2009.

3 Martin A. Makary and Michael Daniel, "Medical Error—the Third Leading Cause of Death in the US," *BMJ* 353 (May 3, 2016): i2139.

4 Donald M. Berwick and Andrew D. Hackbarth, "Eliminating Waste in US Health Care," *JAMA* 307, no. 14 (Apr. 11, 2012): 1513–16

5 Gary Claxton et al., *Employer Health Benefits 2016 Annual Survey*, Henry J. Kaiser Family Foundation, Sept. 14, 2016.

6 Nationally, enrollment in high-deductible health plans has surged nearly 50% over the past two years, rising to nearly 30% of all covered workers. See America's Health Insurance Plans (AHIP), "2015 Census of Health Savings Account—High Deductible Health Plans," Nov. 2015.

7 McKinsey Center for U.S. Health System Reform, "2017 Exchange Market: Plan Type Trends," Nov. 3, 2016.

8 Medicare Payment Advisory Commission (MedPAC), *Report to the Congress: Promoting Greater Efficiency in Medicare*, June 2007.

9 Stephen F. Jencks, Mark V. Williams, and Eric A. Coleman, "Rehospitalizations Among Patients in the Medicare Fee-for-Service Program," *New England Journal of Medicine* 360, no. 14 (Apr. 2, 2009): 1418–28.

10 Employee Benefit Research Institute, "Narrow Provider Networks for Employer Plans," Issue Brief no. 148 (Dec. 14, 2016): "Moreover, once uncertainty about the Cadillac tax is resolved one way or the other, employers may be motivated by ever-increasing health care costs to consider whatever additional changes are still untried that have at least the potential to bend the cost curve without sacrificing essential quality."

11 Ibid. The key distinction between "narrow" and "tiered" networks is that the former outright exclude providers. The latter might keep certain providers in-network but charge higher cost-sharing for using them.

12 J. Matthew Austin et al., "National Hospital Ratings Systems Share Few Common Scores and May Generate Confusion Instead of Clarity," *Health Affairs* 34, no. 3 (Mar. 2015): 423–30.

13 Kenneth J. Arrow, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53, no. 5 (Dec. 1963): 141–49.

14 Zack Cooper et al., "Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms," London School of Economics and Political Science, LSE Health, Working Paper no. 16/2010 (Jan. 2010).

15 William B. Vogt and Robert Town, "How Has Hospital Consolidation Affected the Price and Quality of Hospital Care?" Robert Wood Johnson Foundation, Feb. 2006; and Martin Gaynor and Robert Town, "The Impact of Hospital Consolidation—Update," Robert Wood Johnson Foundation, June 2012.

16 See Federal Reserve Bank of St. Louis, FRED (Federal Reserve Economic Data), "Consumer Price Index for All Urban Consumers: New Vehicles," Mar. 22, 2017.

17 Amitabh Chandra et al., "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review* 106, no. 8 (Aug. 2016): 2110–44.

18 Susannah Fox and Maeve Duggan, "Health Online 2013: One in Three American Adults Have Gone Online to Figure Out a Medical Condition," Pew Research Center, Jan. 15, 2013.

19 Gregory D. Kennedy, Sarah E. Tevis, and K. Craig Kent, "Is There a Relationship Between Patient Satisfaction and Favorable Outcomes?" *Annals of Surgery* 260, no. 4 (Oct. 2014): 592–600.

20 Fadi Hachem, "The Relationships Between HCAHPS Communication and Discharge Satisfaction Items and Hospital Readmissions, *Patient Experience Journal* 1, no. 2 (Fall 2014): 71–77.

21 Jeremy Stoppelman, "Yelp's Consumer Protection Initiative: ProPublica Partnership Brings Medical Info to Yelp," Aug. 5, 2015.

22 Naomi S. Bardach et al., "The Relationship Between Commercial Website Ratings and Traditional Hospital Performance Measures in the USA," *BMJ Quality and Safety* 22, no. 3 (Mar. 2013): 194–202.

23 Benjamin R. Lanard et al., "Yelp Reviews of Hospital Care Can Supplement and Inform Traditional Surveys of the Patient Experience of Care," *Health Affairs* 35, no. 4 (Apr. 2016): 697–705.

24 Aaron E. Carroll, "How Yelp Reviews Can Help Improve Patient Care," *New York Times*, Sept. 12, 2016.

25 For detailed information on these measures, see U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, "National Quality Forum (NQF) Endorsed Individual and Composite Measures," July 12, 2016; and Agency for Healthcare Research and Quality, "Healthcare Cost and Utilization Project (H-CUP)," Overview of Key Readmission Measures, Dec. 12, 2012.

[26] As more payments incorporate patient-satisfaction scores, evidence-based care guidelines may come into conflict with patient expectations. Physicians, for instance, may be under pressures to comply with patient requests for pain medication for chronic back pain rather than prescribe physical therapy. See Shivan J. Mehta, "Patient Satisfaction Reporting and Its Implications for Patient Care," *AMA Journal of Ethics* 17, no. 7 (July 2015): 616–21, at 618: "Given regulations and declining reimbursements, physicians have limited time to spend on each patient visit, and it may be time consuming to explain to patients who expect low-value treatments why they should be withheld. If a physician is faced with penalties for low patient experience scores, it may be the path of least resistance to agree to such requests; even the anticipation of patient dissatisfaction may drive unnecessary or different care in individual cases."

[27] Antoinette Siu, " 'Yelp for Healthcare' Launches with Data on 10,000 California Doctors," *San Francisco Business Times*, Mar. 24, 2017; and California Healthcare Performance Information System, "New Website for Physician Star Ratings," Mar. 22, 2017.

[28] See Health Data NY, "Hospital Inpatient Potentially Preventable Readmissions."

[29] AHRQ Quality Indicators, "Mortality for Selected Procedures, Technical Specifications, Inpatient Quality Indicators 90," May 2013.

[30] AHRQ Quality Indicators, "Mortality for Selected Procedures, Technical Specifications, Inpatient Quality Indicators 91," May 2013.

Abstract

Online tools like ZocDoc, HealthGrades, and Yelp have become popular among people who search for information about physicians and hospitals. Yelp, one of the most widely used platforms, allows patients to rate health-care providers through a five-star rating system that can include narrative text reviews. In 2015, Yelp partnered with ProPublica to publish average wait times, readmission rates, and the quality of communication scores for more than 25,000 hospitals, nursing homes, and dialysis clinics.

According to some research, Yelp reviews correlate with Medicare surveys such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). We don't know, however, how accurate these reviews are for identifying quality doctors and hospitals.

This paper examines whether there is a correlation between Yelp reviews of New York State hospitals and objective measures of hospital quality. We find that higher Yelp ratings are correlated with better-quality hospitals and that they provide a useful, clear, and reliable tool for comparing the quality of different facilities as measured by potentially preventable readmission rates (PPR), a widely accepted metric.

Yelp alone is not, nor can it be, the only guide to quality hospitals. However, when people can choose where they will obtain care, Yelp ratings can provide a helpful guide. These ratings will get even better at helping people as the number of Yelp reviews increases.

New York State policymakers, private and public employers, and Yelp itself thus have an opportunity to help consumers navigate the health-care system by considering the following recommendations:

1. Help make Yelp scores and reviews more visible when consumers are making important decisions about health-care coverage—for instance, when choosing among competing insurers' hospital networks on New York State's health-insurance exchange.

2. Link objective, simple quality metrics onto the Yelp review page for hospitals to allow patients with specific concerns to access more detailed information that would complement and better inform Yelp quality ratings.

3. Fund targeted "hackathons" that find ways to make Yelp and other social media reviews more accessible to high-needs, vulnerable populations—including caregivers for the frail, elderly, non-English-speaking, or low-income minority populations.

By disseminating neutral, clear signals about basic hospital quality, social media tools can also improve the ability of higher-quality hospitals to compete to attract market share, leading to more lives saved and more costs avoided for patients, taxpayers, and employers.

MANHATTAN

INSTITUTE