Disclosing Physician Ratings: Performance Effects and the Difficulty of Altering Ratings Consensus

HENRY EYRING*

ABSTRACT

I examine effects of a health care system's policy to publicly disclose patient ratings of its physicians. I find evidence that this policy leads to performance improvement by the disclosed, subjective ratings and also by undisclosed, objective measures of quality. These effects are consistent with multitasking theory, in that physicians respond to the disclosure by providing more of a shared input—time with patients—that benefits performance by ratings and underlying quality. I also find, as predicted by information cascade theory, that the

ratings become jammed to some degree near initially disclosed values. Specifically, raters observe the pattern of initial ratings and follow suit by providing similar ratings. Finally, I find evidence that physicians anticipate rating jamming and so concentrate their effort on earlier performance in order to set a pattern of high ratings that later ratings follow. These results demonstrate that the disclosure of subjective ratings can benefit performance broadly but can also shift effort toward earlier performance.

JEL codes: D83, D90, I10, I11, L15

Keywords: disclosure; real effects; information cascade; multitasking

1. Introduction

I examine effects of a health care system's policy to publicly disclose patient ratings of its physicians. Many organizations, in settings such as retail, education, and transportation, are similarly disclosing subjective ratings of their performance.[1] High ratings can attract revenue when they are disclosed (Chevalier and Mayzlin [2006], Hanauer et al. [2014]), and this may reward high performance in a way that motivates effort provision and affects underlying quality. Although economic and accounting research has documented performance effects that result from various disclosures (Jin and Leslie [2003], Christensen et al. [2017]), this literature has not yet considered dynamics that are at play when disclosure reveals subjective ratings. Economic theory suggests that these dynamics may affect performance across subjective and objective measures, and may also create incentives to shift effort over time. I address these ideas empirically using data from a health care system.

My study's field site, University of Utah Health Care (UUHC), implemented a policy in 2012 to disclose ratings for each of its physicians who had received at least 30 ratings in the preceding 12 months.[2] I draw on detailed, patient-visit-level data on multiple measures of performance for physicians, who were either included in or excluded from disclosure based on the number of ratings they had previously received. Using difference-in-differences (DiD) analysis, I compare changes around the time of disclosure for these groups of physicians. My models control for flexible time trends, static differences among physicians, and an extensive set of patient and visit characteristics. I use this identification strategy to provide initial evidence on performance effects of the spreading disclosure of subjective ratings.

---

[1] For example, eBay discloses ratings of its product sellers (eBay [2020]), Uber discloses ratings of its rideshare drivers (Uber [2020]), the city of San Francisco discloses ratings of its local government (City Performance Team [2019]), and the state of Texas requires that its state-run universities disclose ratings of their faculty members (Texas Tech University [2020]).

[2] "Physician ratings" is one of the popular terms used to describe patient ratings of physicians (Glover [2014]), and I adopt this terminology.

Research on disclosure's performance effects has shown that, when performance by disclosed measures attracts revenue or investment, this produces reputational incentives that can lead to improvement by disclosed measures (Jin and Leslie [2003], Christensen et al. [2017]). I find that the disclosure of physician ratings produces similar reputational incentives in my setting. Specifically, I show that rating disclosure directs patients to higher-rated physicians and that this leads to a roughly 12% increase in pay for these physicians. Although these incentives would plausibly lead to improved performance by ratings, it is not theoretically straightforward how this will affect performance by objective measures of quality. On one hand, some practitioners and scholars have predicted that effort to improve performance by ratings may divert effort away from, and thereby harm, performance by objective measures (Friedberg, Safran, and Schneider [2012], Bond [2015]). On the other hand, multitasking theory on input-sharing suggests that effort to improve performance by ratings could work through shared inputs to boost performance by objective measures (Feltham and Xie [1994], Mullen, Frank, and Rosenthal [2010]).

Motivated by this tension, I test how rating disclosure affects performance by the disclosed, subjective ratings and by undisclosed, objective measures of quality. Health economics research suggests that physician ratings and objective measures of medical quality share some inputs, including the time that a physician spends with a patient.[3] Using data from computer timestamps during visits, I estimate that rating disclosure leads physicians to spend roughly 25% more time with each patient.[4] I then show that rating disclosure works partly through this increase in physician time with patients to improve performance by subjective ratings and objective quality measures.[5] These performance effects are enough to lift a physician with median ratings toward the top of the predisclosure rating distribution and to boost performance on objective quality measures by roughly 37%.

---

[3] Friedberg, Safran, and Schneider [2012] provide a survey of literature that demonstrates a positive relationship between physician ratings and objective measures of medical quality, Lin et al. [2001] document a positive relationship between physician time spent with patients and physician ratings, and Neprash [2016] documents a positive relationship between physician time spent with patients and objective measures of medical quality.

[4] See Hribar et al. [2015] for evidence validating the use of timestamps from electronic health records to measure physician time spent with patients.

[5] In a supplemental analysis, I examine whether the increase in physician time with patients occurs through a decrease in patient volume. I find instead that the increase in physician time with patients occurs along with an increase in visit volume. This is plausible in light of survey data that suggest that physicians spend the majority of their time in the office away from patients (Sinsky et al. [2016]), time that can be redirected to patients through better office management. In line with that explanation for the increase in physician time with patients, physicians at UUHC reported increasing time with patients by delegating clerical work and using electronic notifications to keep better track of when patients were waiting for the physician to enter an examination room.

In addition to testing for effects across performance measures, I explore whether rating disclosure creates dynamic incentives, or incentives to shift effort over time (Casas-Arce and Martínez-Jerez [2009], Bouwens and Kroos [2011]). These incentives may be present if physicians expect that raters will, in line with principles of information cascade, observe early disclosed ratings and follow suit by providing similar ratings (Welch [1992], Anderson and Holt [1997]). Physicians who anticipate an information cascade, or herd, in their ratings may perform best initially in order to set a pattern of high ratings that later ratings follow.

To document herding in disclosed ratings, I show that, after rating disclosure, a physician's ratings tighten around his or her ratings as they stood at the time of disclosure. Additionally, I show that ratings become less responsive to changes in underlying quality measures after the disclosure, which is consistent with ratings jamming near early disclosed values. I then offer evidence that herding in disclosed ratings generates dynamic incentives. Specifically, I find that physicians perform best leading up to and shortly after the disclosure. This is more pronounced for physicians who, due to the large volume of Web traffic and ratings they receive, would plausibly anticipate a stronger effect of herding on their ratings.

This paper's primary contribution is to research on disclosure's real effects, or effects on the behavior of the disclosing entity (Leuz and Wysocki [2016], Christensen, Floyd, and Maffett [2020]). Policy makers and businesses are increasingly using disclosure as a tool to motivate performance improvement (The Economist [2014], Cannizzaro and Weiner [2016]), and real effects research has examined whether and how such performance effects occur (Jin and Leslie [2003], Christensen et al. [2017]). This literature has yet to consider performance effects of disclosure that reveals subjective ratings, a practice that is spreading among governments, nonprofits, and companies. I demonstrate that disclosing subjective ratings can: (1) lead to better performance by subjective and objective measures, and (2) create incentives to shift effort toward early performance.

My results may also inform research in other areas of economics and accounting. First, I offer evidence to support the prediction from recent multitasking models that incentives for one measure can benefit performance by other measures (Mullen, Frank, and Rosenthal [2010]). This helps to address the large gap that has formed as multitasking theory has significantly outpaced empirical research (Hong et al. [2013]), and sheds light on the way that partial incentives can create broad performance benefits in jobs that involve responsibility for many measures (Lindbeck and Snower [2000]). Second, I document an information cascade in subjective performance evaluations (Welch [1992], Anderson and Holt [1997]). This extends accounting research that explores how subjective performance evaluations form (Bol et al. [2010], Bol [2011]). Finally, I offer evidence that an information cascade can create incentives for the evaluated agent

to shift effort toward earlier performance. Future research could examine whether similar dynamic incentives arise in other contexts where information cascade has been studied, such as in political polls and in IPOs (Bikhchandani, Hirshleifer, and Welch [1992], Welch [1992], Eyster and Rabin [2010]).

2. Theory and Hypotheses

A growing literature on disclosure's real effects examines how disclosure affects the behavior of the disclosing party (Leuz and Wysocki [2016], Christensen et al. [2017], Granja [2018]). Studies have shown that disclosure can boost performance by disclosed measures in contexts ranging from mine safety to restaurant hygiene.[6] As in those studies, the disclosure in my setting plausibly creates incentives to improve performance by disclosed measures. In particular, I document that rating disclosure directs the flow of patients to higher-rated physicians. Given that physicians at UUHC are paid partly based on the volume of visits and services they provide, rating disclosure's effect on patient choice of physicians creates implicit financial incentives for physicians to achieve high ratings. I discuss these incentives in greater detail in section 5.

To understand rating disclosure's effects across subjective and objective measures, I draw on multitasking theory. A foundational model in Holmstrom and Milgrom [1991] explains that incentives for one measure increase the opportunity cost of providing effort toward a second measure and so direct effort away from the second measure. In the context of rating disclosure, this model implies that physician effort to improve performance by subjective ratings may divert effort away from and so harm performance by objective quality measures. However, the Holmstrom and Milgrom [1991] model assumes that each performance measure is a function of a single, unique input, which the authors note is not likely to hold in many settings. Extensions to Holmstrom and Milgrom [1991] relax this assumption and explain how, when measures share inputs, incentives to perform better by one measure may lead to better performance by another (Feltham and Xie [1994], Mullen, Frank, and Rosenthal [2010]).

There are a few reasons to expect that the disclosure of physician ratings could, in line with multitasking theory on input-sharing, boost performance by objective measures. First, a number of studies suggest that physician ratings and objective measures of medical quality share inputs. These inputs include clear communication, accurate diagnosis, and physician time spent with patients (Friedberg, Safran, and Schneider [2012],

---

[6] See Jin and Leslie [2003], Beyer and Guttman [2012], and Christensen et al. [2017] for examples of research on real effects. For reviews of disclosure literature, see Healy and Palepu [2001] regarding disclosure in capital markets, and Leuz and Wysocki [2016] regarding a range of financial and nonfinancial disclosures and their effects on information users and disclosers.

Hachem et al. [2014], Neprash [2016]). Second, physicians at UUHC reported that rating disclosure led them to provide more of some of these inputs, including time spent with patients. For example, some physicians reported delegating clerical tasks and spending more of their time in examination rooms with patients. Other physicians reported placing timers and lights outside of examination rooms to more quickly identify and begin visits with awaiting patients.[7]

H1 predicts that disclosing physician ratings leads to performance improvement by ratings. H2 predicts that this will yield a positive performance spillover to performance by objective quality measures. H3a and H3b predict that physician time spent with patients, which UUHC measures using computer timestamps, will mediate the effect of the disclosure on ratings and objective quality measures.

Although these hypotheses relate primarily to research on disclosure's real effects, they also help to reduce the gap between multitasking theory and empirical work. Consonant with the observation that "theory related to multitasking is decades ahead of the empirical evidence" (Hong et al. [2013]), studies in this stream have not yet offered evidence to establish a performance spillover through the input-sharing mechanism predicted by Mullen, Frank, and Rosenthal [2010]. H1–H3b outline a mediation analysis to test for this result.

H1: Disclosing physician ratings leads to improvement by physician ratings.

H2: Disclosing physician ratings leads to improvement by objective quality measures.

H3a: Physician time spent with patients mediates rating disclosure's effect on ratings.

H3b: Physician time spent with patients mediates rating disclosure's effect on objective quality measures.

To predict how the disclosure will affect performance over time, I draw on information cascade literature. Information cascade models explain that a "herd" can form through Bayesian updating (Banerjee [1992], Anderson and Holt [1997]). In a herd, an agent observes a publicly visible signal and follows suit by providing a similar signal even when this conflicts with his or her private information. Analytical models explain that a

---

[7] Although rating disclosure could lead physicians to provide inputs that ratings and objective quality measures share, it could also lead physicians to take actions that produce superficial improvement in ratings. Research has demonstrated this type of superficial improvement following disclosure. For example, Christensen, Floyd, and Maffett [2020] find that the disclosure of hospitals' charges leads hospitals to reduce charges without reducing the payments they receive, which is possible given that hospitals can decouple charges from payments. Dranove et al. [2003] show that cardiologists respond to mortality rate disclosure by avoiding surgical paths for sicker patients who are less likely to survive an operation. In my setting, physicians may try to superficially improve their ratings by, for example, smiling more or being friendlier without making changes that would also benefit medical quality.

herd can form quickly and lead to a large divergence between the consensus opinion about a subject and its underlying qualities or characteristics (Welch [1992]). If physicians anticipate that their ratings will form through a herd following disclosure, this could incentivize physicians to focus effort on early performance in order to start a herd in a desired direction.

Research has demonstrated herding in the lab when there are incentives to accurately estimate an unknown value (Anderson and Holt [1997]). However, it is less clear whether herding will occur with subjective ratings, where there is no objectively accurate estimation and there is no financial incentive for accuracy. Theory from information systems research suggests that raters who disagree with posted ratings may even exaggerate their ratings in the direction of disagreement, trying to sway the average toward their disparate views (Eryarsoy and Piramuthu [2014]).

Empirical studies have yet to provide consistent evidence of herding in ratings as predicted by information cascade models. A lab study finds that some raters move their ratings toward and other raters move their ratings away from a consensus after the consensus is disclosed, but does not find a net effect on the average rating (Eryarsoy and Piramuthu [2014]). A field experiment finds a net effect of arbitrarily assigning a "thumbs-up" rating, but not a "thumbs-down" rating, on the average rating (Muchnik, Aral, and Taylor [2013]). The lab study notes that it is limited by a small sample size, and the field study notes that its contrasting results may stem from its design, which does not have a counterfactual set of ratings that are kept private from subsequent raters. My setting offers a large sample size and a counterfactual in the form of physicians who were excluded from rating disclosure. These features help overcome challenges faced by prior empirical research on information cascade in ratings.[8]

I use two independent and mutually reinforcing methodologies to examine whether rating disclosure introduces information cascade in ratings. The first methodology follows prior empirical research on information cascade. In particular, I look for a tighter distribution of signals, herein ratings, around a prior average signal, herein a physician's prior average rating, when that average is disclosed than when it is not (Anderson and Holt [1997], Eryarsoy and Piramuthu [2014]). In my second test of herding, I exploit the availability of data on subjective ratings and on objective measures to test the idea that herding will cause a divergence between an opinion-based signal and underlying quality (Welch [1992]). Specifically, I test whether ratings become less sensitive to changes in objective measures of quality after rating disclosure. H4a and H4b lay out the predictions for

---

[8] Muchnik, Aral, and Taylor [2013], Eryarsoy and Piramuthu [2014], and my study examine anonymous raters. A related set of studies considers how, when raters are identifiable, they may conform in their opinions in order to send a signal about similar likes or dislikes and thereby gain social acceptance (Schlosser [2005], Lee, Hosanagar, and Tan [2015]). Herding through Bayesian updating, which is the mechanism in information cascade models, can occur whether or not raters are identifiable. My study fits within literature that considers this mechanism.

these two complementary tests of herding. After testing for the presence of herding in ratings, I explore how the prospect of herding in ratings may create incentives for physicians to shift effort over time. H5 predicts that performance effects of disclosure will be strongest initially, which would occur if physicians attempt to set a favorable pattern of early ratings.

By advancing these hypotheses, I extend research on information cascade and on subjective performance evaluation in two ways. First, research on information cascade has described how herds form, but has not previously considered how an evaluated agent might optimally behave to influence the direction of a rating herd (Welch [1992], Anderson and Holt [1997], Eyster and Rabin [2010]). I test for this behavior by exploring whether rating disclosure, which introduces the prospect of a herd forming, leads physicians to shift effort toward earlier performance. Second, research on subjective performance evaluation has examined how evaluations form, including what causes them to diverge from underlying performance (Bol et al. [2010], Bol [2011]). I add to this research by documenting a divergence between subjective evaluations and objective measures of quality that occurs when prior evaluations are visible to later evaluators.

> *H4a*: Disclosing physician ratings leads to a tighter distribution of ratings around each physician's rating consensus.
> *H4b*: Disclosing physician ratings leads ratings to change less with changes in underlying quality.
> *H5*: Disclosing physician ratings leads to the greatest improvement in objectively measured quality soon after the disclosure's announcement.

After testing H4a–H5, I conduct cross-sectional tests to address the alternative explanation that, when physicians focus their effort on initial performance, this is due to the novelty of disclosure and not due to the prospect of an information cascade. If physicians are shifting effort out of concern for the formation of a herd, then this behavior should be more pronounced when the physician expects that the disclosed ratings will receive more public attention and consist of a larger and so more persuasive sample size, which are conditions that promote the formation of a herd (Welch [1992]). Using data on Web traffic to physician Web pages and on rating volume prior to the disclosure, I test for more pronounced shifts in physician effort toward initial performance under these conditions.[9]

---

[9] Throughout the sample period, administrators sent physicians periodic reports of their ratings and of Web traffic to their official Web pages, where ratings were disclosed. Thus, physicians were made aware of the amount of Web traffic and number of ratings they were receiving.

3. Setting

3.1 FIELD SITE AND DECISION TO DISCLOSE RATINGS

This paper's field research site is UUHC, an academic medical system that has four hospitals, 12 community clinics, and several specialty centers. The system receives over 1 million outpatient visits and 30,000 inpatient visits annually, offering services ranging from primary care to the most advanced types of cancer treatment.

UUHC uses a third-party survey vendor, Press Ganey Inc., to distribute patient satisfaction surveys. Press Ganey serves over half of the nation's hospitals and automatically emails a patient satisfaction survey following each patient visit at UUHC. The survey asks a patient to rate a physician on several criteria, listed in appendix A.

Administrative discussions of the possibility of disclosing these ratings arose for a number of reasons. One was to create incentives for physicians to improve their ratings. By making physician ratings visible to patients who could then use the ratings to select among physicians, the disclosure could create reputational incentives for physicians to achieve high ratings. Another factor contributing to the decision to disclose physician ratings was the health care system's goal to be a visible leader of health care quality and transparency. As expected, the decision to disclose ratings led to substantial press coverage in international media (e.g., Lee [2014], The Economist [2014]). Finally, administrators and physicians referenced their underlying belief that it is important to help patients choose among physicians, and that patients should be able to learn from each other in that process.

3.2 RATING DISCLOSURE

Administrators chose to use physicians' official online Web pages, which were already visible to the public, as the venue for disclosing ratings. A system-wide email in November 2012 announced the rating disclosure and ratings were first disclosed at the top of physicians' official online Web pages in December 2012.[10] The only criterion for a physician's inclusion in disclosure was that he or she had received 30 or more ratings in the 12 months preceding a rating disclosure.

Physicians who met that criterion in December 2012 were the first to have their ratings disclosed, and I term these "First-Disclosed Physicians." In July 2013, any physicians who failed to meet the criterion in December 2012 but met the criterion in July 2013 had their ratings disclosed. I term these "Second-Disclosed Physicians." I term physicians who failed to meet the criterion for both disclosures during my sample period "Nondisclosed Physicians." A physician's disclosed rating was the physician's prior 12-month average as it stood at the most recent disclosure event that the

---

[10] Along with quantitative ratings, UUHC posted all comments regarding the physician that did not identify the patient or contain slander or profanity. Appendix B contains example comments.

Physician-Rating Disclosure Timeline

Disclosure Announced — System-wide email announced the rating disclosure to physicians

Disclosure 1 — Average rating disclosed for any physician with at least 30 ratings in the prior 12 months

Disclosure 1 Update — Average rating updated for physicians included in Disclosure 1 and with at least 30 ratings in the prior 12 months

Disclosure 2 — Average rating disclosed for physicians not included in Disclosure 1 and with at least 30 ratings in the prior 12 months
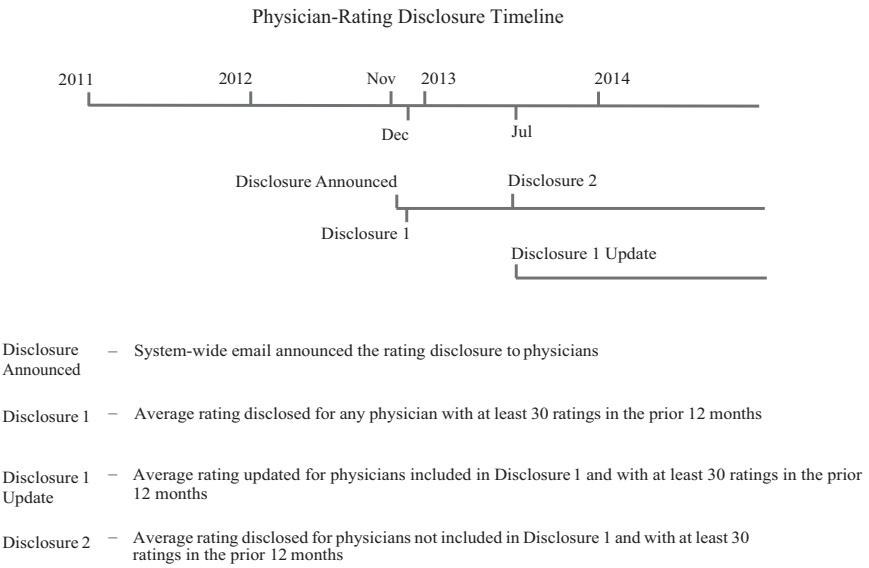
FIG. 1.—Physician-rating disclosure timeline. This figure shows the timeline of physician-rating disclosure. I term physicians included in Disclosure 1 "First-Disclosed Physicians." I term physicians excluded from Disclosure 1 but included in Disclosure 2 "Second-Disclosed Physicians." I term physicians excluded from both disclosures "Nondisclosed Physicians."

FIG. 2.—Example physician online profile. This figure shows an example physician online profile after physician-rating disclosure. The consensus rating is shown in the top right corner. The number of ratings that the consensus rating consists of is shown immediately below. Below that number of ratings is a link to patient comments regarding the physician.

physician met the survey-count criterion for. Once disclosed, a physician's consensus rating stayed constant on the physician's Web page until the next disclosure the physician qualified for. Figure 1 details the disclosure timeline, and figure 2 shows an example physician online profile after rating disclosure.

I interviewed UUHC's administrators tasked with gathering and disclosing physician ratings to learn about each of these three groups.

Two-hundred and seventy-three physicians—about two-thirds of my sample—were included in the first disclosure. Administrators explained to me that the other roughly one-third of physicians would have either: (1) too recently joined UUHC to have accumulated enough ratings until the second disclosure (i.e., making these physicians "Second-Disclosed Physicians"), or (2) have consistently received fewer than 30 ratings a year (i.e., making these physicians "Nondisclosed Physicians").

I find that Second-Disclosed Physicians are few (22) and joined UUHC about five years later on average than First-Disclosed or Nondisclosed Physicians. In subsection 5.7, I discuss robustness tests to control for the recency of physicians joining UUHC. My results are robust to excluding Second-Disclosed Physicians from my analysis and to matching on the length of time that a physician has been with UUHC.

Roughly one-fourth of the physicians in my sample are Nondisclosed Physicians. Administrators noted a few reasons why these 93 physicians would consistently receive fewer than 30 ratings a year. First, UUHC physicians teach and conduct research through the University of Utah Medical School. Tenure-track physicians spend more of their time on these activities and so receive fewer survey responses. Second, some specialties involve fewer patient visits and have a larger percentage of their doctors in the Nondisclosed Physician group. In section 5, I discuss robustness tests to control for these differences among the samples of physicians that I compare. This includes matching so that there are equal numbers of tenure-track physicians and physicians from each specialty in each sample.

4. Data

My tests draw on proprietary data spanning from 2011 to late 2014. Table 1 contains descriptive statistics for physicians, patients, and visits. Appendix C contains variable definitions.

4.1 PHYSICIAN, PATIENT, AND VISIT CHARACTERISTICS

Physician data include gender, education, age, number of years employed by UUHC, and whether the physician is on a tenure track at the University of Utah Medical School. Patient data include gender, age (winsorized above age 89 in compliance with privacy standards), and whether the patient speaks English.[11]

Data on visit characteristics include the insurer and charges for the visit, whether the visit was the patient's first to the physician, and two measures

---

[11] I incorporate patient age as indicator variables to account for nonlinearity in the relationship between age and my dependent variables. In the analyses of ratings, herding, and weekly *RVUs* and visits, I use indicators from psychology research that represent differences in emotion and cognition that would plausibly influence ratings and patient responses to them (Newman and Newman [2014]). In the analyses of objectively measured quality and physician time with patients, I use indicators as outlined by the International Epidemiological Association (IEA) [2019] for research on physical health.

TABLE 1

Sample Selection and Descriptive Statistics

Panel A: Sample selection	
Physician-week visits and RVUs	
Initial observations	118,774
Exclude physicians who exit the sample before or enter the sample after disclosure announcement	(47,357)
Sample for physician-week visits and RVUs	71,417
Ratings	
Initial observations	178,334
Exclude physicians who exit the sample before or enter the sample after disclosure announcement	(77,968)
Sample for ratings	100,366
Exclude physicians who enter the sample more recently than a year before first disclosure	(5,948)
Restrict sample to one year before and one year after first disclosure	(34,494)
Sample for absolute difference	59,924
Procedures	
Initial observations	48,839
Exclude physicians who exit the sample before or enter the sample after disclosure announcement	(7,245)
Sample for quality deductions	41,594
Time with patient	
Initial observations	296,033
Restrict sample to visits entered in scheduling system as a regular initial or follow-up visit	(116,222)
Exclude physicians who exit the sample before or enter the sample after disclosure announcement	(48,886)
Sample for time with patient	130,925
	(Continued)

TABLE 1
Continued

Panel B: Physician-week visits and RVUs descriptive statistics

Unit of Observation	First-Disclosed Physician			Second-Disclosed Physician			Nondisclosed Physician		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Physician									
Gender	273	0.36	0.48	22	0.36	0.48	93	0.34	0.47
MD	273	0.79	0.40	22	0.81	0.38	93	0.84	0.36
Age	273	48.96	10.19	22	45.77	8.03	93	47.13	8.57
Years With UUHC	273	9.85	4.96	22	4.79	3.59	93	8.23	4.05
Tenure Track	273	0.31	0.46	22	0.22	0.41	93	0.37	0.48
Physician-week									
RVUs Per Week	52,952	121.76	124.37	4,457	106.77	123.07	14,008	30.59	72.71
Visits Per Week	52,952	39.10	34.96	4,457	19.44	21.28	14,008	9.16	10.13
Gender	52,952	0.57	0.49	4,457	0.55	0.49	14,008	0.57	0.49
Age	52,952	50.02	22.35	4,457	55.31	20.30	14,008	50.43	22.24
CCI	52,952	0.18	0.42	4,457	0.21	0.41	14,008	0.19	0.42
Medicare or Medicaid	52,952	0.25	0.13	4,457	0.27	0.13	14,008	0.22	0.18

(Continued)

TABLE 1
Continued

Panel C: Ratings descriptive statistics

Unit of Observation	First-Disclosed Physician			Second-Disclosed Physician			Nondisclosed Physician		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Physician									
Gender	273	0.36	0.48	22	0.36	0.48	93	0.34	0.47
MD	273	0.79	0.40	22	0.81	0.38	93	0.84	0.36
Age	273	48.96	10.19	22	45.77	8.03	93	47.13	8.57
Years With UUHC	273	9.85	4.96	22	4.79	3.59	93	8.23	4.05
Tenure Track	273	0.31	0.46	22	0.22	0.41	93	0.37	0.48
Patient visit									
Gender	94,948	0.60	0.48	2,801	0.64	0.47	2,617	0.49	0.50
Age	94,948	49.66	19.86	2,801	55.49	17.76	2,617	46.88	23.92
English Speaking	94,948	0.98	0.13	2,801	0.98	0.11	2,617	0.98	0.10
Charges ($)	94,948	282	1,182	2,801	437	1,719	2,617	313	872
RVUs	94,948	1.94	2.95	2,801	2.00	2.61	2,617	2.14	8.57
CCI	94,948	0.01	0.14	2,801	0.00	0.10	2,617	0.01	0.14
Medicare or Medicaid	94,948	0.17	0.37	2,801	0.22	0.42	2,617	0.27	0.44
First Visit	94,948	0.25	0.43	2,801	0.29	0.45	2,617	0.17	0.38
Rating	94,948	4.70	0.52	2,801	4.71	0.51	2,617	4.74	0.48
Absolute Difference	57,969	0.35	0.36	881	0.37	0.42	1,074	0.29	0.33

(Continued)

TABLE 1
Continued

Panel D: Procedures descriptive statistics

Unit of Observation	First-Disclosed Physician			Second-Disclosed Physician			Nondisclosed Physician		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Physician									
Gender	139	0.28	0.44	13	0.38	0.34	36	0.31	0.46
MD	139	0.94	0.23	13	1.00	0.00	36	0.86	0.27
Age	139	48.22	8.52	13	44.03	7.07	36	43.05	6.74
Years With UUHC	139	9.31	3.96	13	3.53	1.61	36	6.55	3.89
Tenure Track	139	0.57	0.49	13	0.23	0.43	36	0.27	0.37
Patient visit									
Gender	35,572	0.54	0.49	1,242	0.50	0.50	4,780	0.55	0.49
Age	35,572	49.28	21.56	1,242	49.87	18.78	4,780	43.80	23.66
Charges ($)	35,572	44,622	102,770	1,242	32,239	60,482	4,780	44,957	139,024
RVUs	35,572	2.14	3.41	1,242	1.61	2.77	4,780	2.04	3.38
CCI	35,572	0.12	0.46	1,242	0.08	0.39	4,780	0.25	0.72
Medicare or Medicaid	35,572	0.04	0.21	1,242	0.02	0.17	4,780	0.09	0.29
Quality Deductions	35,572	0.04	0.20	1,242	0.03	0.17	4,780	0.04	0.20

(Continued)

TABLE 1
Continued

Panel E: Time with patient descriptive statistics

Unit of Observation	First-Disclosed Physician			Second-Disclosed Physician			Nondisclosed Physician		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Physician									
Gender	250	0.29	0.45	22	0.36	0.48	65	0.43	0.49
MD	250	0.78	0.41	22	0.81	0.38	65	0.50	0.50
Age	201	48.62	10.22	22	45.77	8.03	52	43.84	7.42
Years With UUHC	250	10.71	5.61	22	4.79	3.59	65	8.10	4.62
Tenure Track	250	0.36	0.48	22	0.22	0.41	65	0.20	0.40
Patient visit									
Gender	122,209	0.60	0.48	4,210	0.58	0.49	4,506	0.46	0.49
Age	122,209	44.23	24.17	4,210	43.89	22.25	4,506	55.23	18.83
Charges ($)	122,209	177	405	4,210	209	566	4,506	262	1,434
RVUs	122,209	2.97	4.91	4,210	4.29	6.31	4,506	5.87	13.62
CCI	122,209	0.01	0.16	4,210	0.01	0.18	4,506	0.06	0.35
Medicare or Medicaid	122,209	0.46	0.49	4,210	0.42	0.49	4,506	0.47	0.49
Time With Patient	122,209	16.77	17.01	4,210	24.05	22.73	4,506	27.59	24.95

Panel A presents the sample selection for tests regarding physician-week visits and *RVUs*, physician ratings, procedures, and physician time with patients. Panel B presents descriptive statistics for tests regarding physician-week visits and *RVUs*. Panels C and D present descriptive statistics for tests regarding physician ratings, and for tests regarding procedures, respectively. Panel E presents descriptive statistics for tests regarding the amount of time in minutes that a physician spends with a patient during a visit. In the data on physician time spent with patients, I have physician ages for 80% of the First-Disclosed and Nondisclosed physicians. Physician fixed effects in my models subsume physician characteristics including physician ages. Physician ages are as of January 1, 2015, and patient ages as of the time of the visit and treated as 90 if above 89 in compliance with privacy standards.

of severity and complexity of the patient's condition. One is the number of relative value units (*RVUs*) produced during the visit, a number that reflects the severity of the case and the related complexity of treatment. The amount of Medicare reimbursement for a visit rises proportionally to *RVUs*. The second is the Charlson Comorbidity Index (*CCI*), a weighted score that represents the disease burden of the patient.[12] When the patient has a comorbid condition, the *CCI* takes a value of 1, 2, 3, or 6, in general proportion to the likelihood of mortality within one year associated with the comorbid condition (e.g., ulcers, diabetes).[13]

4.2 TIME WITH PATIENT

For a subset of visits, I have data on the amount of time that a physician spent with a patient during the visit, or *Time With Patient*. UUHC collects these data using timestamps from the processing of a patient's electronic health record during successive stages of a patient visit. UUHC and other health care systems review time data gathered in this manner for operational analysis (Danciu et al. [2014]). Health care methodological research has validated this approach to time measurement by comparing these computer-generated times to estimates that are based on observation by a human (Hribar et al. [2015]). I restrict my analysis of *Time With Patient* to visits that the health care system designates regular initial or follow-up visits. This helps my models to capture increased *Time With Patient* that would come from a physician being available for more of a given visit, rather than from a change in visit purposes that can lead a centralized scheduler to allocate more or less time for the visit. Internal researchers gathered time data as far back in time as they deemed possible for each of UUHC's clinics. I control for clinic fixed effects in my analysis of time data so that my results are not attributable to static differences across clinics.

4.3 OBJECTIVE QUALITY MEASURES

I measure objective quality improvement as a decrease in *Quality Deductions*. This is a categorical variable that is the sum of instances of flaws in care. These flaws are failure to meet process standards (e.g., failure to provide an advised medication prior to a procedure), instances of a hospital-acquired condition (e.g., an infection from a medical instrument), and

---

[12] See Sundararajan et al. [2004] for a description of the *CCI*, and Dafny [2005], Chandra, Gruber, and McKnight [2010], and Doyle Jr. [2011] for examples of its use in research.

[13] The conditions are recorded at the time of a procedure. Analyses of objectively measured quality regard procedures and are thus able to include *CCI* as measured at the given visit. Ratings and time with patient are measured for visits regardless of whether the visit included a procedure, and thereby regardless of whether *CCI* is measured. For analyses of ratings and time with patient, I include *CCI* as its value for the patient in UUHC visits during the six-month window centered at the rated visit. The results are robust to narrowing this window to three months or expanding it to one year. For analyses of weekly *RVUs* or visits, I include *CCI* as measured on average for the physician during the given week.

readmissions to the emergency department within three months of discharge. Health care regulators and researchers commonly use these occurrences as proxies for avoidable failures in quality (Joynt, Orav, and Jha [2011], Andel et al. [2012]).

4.4 PHYSICIAN RATINGS

Patients can rate their physician after every visit by responding to an automated patient satisfaction survey sent by email. The survey content and distribution are not subject to the discretion of the physician. Patients answer the questions on a Likert scale of 1–5 with 1 indicating "very poor," and 5 indicating "very good." *Rating* is a physician's average rating for a visit. *Rating* components are listed in appendix A. When a physician was included in a disclosure, his or her 12-month average of *Rating* prior to that disclosure was displayed atop his or her official Web page.

Rating is generally very high, with an average in the sample for this study of roughly 4.7 of 5. This raises the possibility that institutional factors specific to UUHC lead only satisfied patients to respond to surveys, but the high average *Rating* is representative of hospitals nationally. For UUHC's peer group of 120 academic hospitals that use Press Ganey surveys, the average is roughly 4.6 of 5.

5. Analysis

5.1 IDENTIFICATION STRATEGY OVERVIEW

My main analyses use DiD models that compare changes around the time of disclosure for physicians who were included in disclosure and those who were excluded. I use physician fixed effects to control for static differences among physicians and time fixed effects to control for common time trends. In models that use this structure to test for a treatment effect, an indicator variable set to equal 1 for treated individuals in the posttreatment period captures the estimated effect.[14]

In tests of the disclosure's performance effects, my models consider physicians whose ratings were disclosed at any point during my sample (i.e., First-Disclosed and Second-Disclosed Physicians) as treated at the time of the disclosure's announcement. Administrators explained that physicians who were included in disclosure during my sample would likely have anticipated, at the time of the disclosure's announcement, that the ratings they were receiving at that time would eventually be disclosed. These physicians would plausibly have responded with efforts to improve performance from that point forward.

For tests of patient responses to disclosure, my models consider physicians whose ratings were disclosed at any point during my sample (again,

---

[14] See Duflo [2002], Dranove et al. [2003], and Acemoglu, Hassan, and Tahoun [2018] for descriptions of similarly specified DiD models with individual and time fixed effects.

First-Disclosed and Second-Disclosed Physicians) as treated at the time when the given physician's ratings were disclosed, since patients could not have responded to the physician's ratings until they were disclosed. Thus, to measure effects of the disclosure on patient demand and on herding, I use the date at which a physician was first included in disclosure as the time of treatment. This was December 2012 for First-Disclosed Physicians and July 2013 for Second-Disclosed Physicians.

The key assumption for DiD identification is that, absent treatment, there would have been parallel trends for the treated and nontreated groups. This assumption is unlikely to hold if there were dissimilar preperiod trends or if there were contemporaneous changes, unrelated to disclosure, that caused a divergence in trends. In subsection 5.7, I apply robustness tests that prior research has outlined to help resolve these concerns (Duflo [2002], Christensen et al. [2017]).

5.2 DISCLOSURE-RELATED INCENTIVES

I first examine physicians' reputational concerns following the disclosure. UUHC adjusts physician pay based on the number of *RVUs* that a physician accumulates through conducting visits. Each visit generates *RVUs*, and more complex visits generate more *RVUs*. Administrators did not provide research access to physician-contract terms but explained that it would be typical for 50% or more of a physician's pay to rise in proportion to *RVUs*.

In interviews, physicians mentioned their expectation that rating disclosure directs patients to higher-rated physicians, and thereby generates more *RVUs* and related pay for these physicians. I test for rating disclosure's effects on visit volume and revenue using the following models:

$$Visits\ Per\ Week_{pytw} = \alpha + \delta Physician_p + \lambda Year_y + \omega Period_t + \varsigma Controls_w$$
$$+ \beta Disclosed_{pt} + \varepsilon_{pytw}, \tag{1}$$

$$RVUs\ Per\ Week_{pytw} = \alpha + \delta Physician_p + \lambda Year_y + \omega Period_t + \varsigma Controls_w$$
$$+ \beta Disclosed_{pt} + \varepsilon_{pytw}, \tag{2}$$

where *Visits Per Week* is the number of visits conducted by physician p in week w, *RVUs Per Week* is the number of *RVUs* produced by physician p in week w, y indexes years, t indexes time periods segmented by disclosure events, and *Disclosed* is an indicator equal to 1 in the time period following which the given physician's ratings were disclosed, if ever during my sample.

Physician fixed effects control for physician characteristics, including membership in or exclusion from the treated group. Year fixed effects control for time trends that are common among physicians, and period fixed effects control for changes in these common time trends around disclosure events within a calendar year. β captures the DiD estimation of the effect of rating disclosure on the dependent variable. The controls vector consists

of a physician-week's average for each of the following: whether the visit's insurer was Medicare or Medicaid, *CCI*, and patient gender and age. My estimates of these models and models 3–5 cluster standard errors at the specialty level.[15] This adjusts for correlation of dependent variables within services.[16]

I estimate models 1 and 2 for physicians who had above-median ratings and then again for physicians who had median or below-median ratings at the beginning of the disclosure regime. In my analysis of *RVUs Per Week*, I winsorize this variable at the 90th percentile so that my results are not explained by outlier patient procedures that have extreme prices and associated *RVUs*.[17] Table 2 shows the results of estimating models 1 and 2. I find that rating disclosure leads to a change in visit volume and *RVUs* that is generally proportional to the physician's rating. Specifically, I estimate that disclosure leads to 2.4 more visits and 11.54 more *RVUs* per week for physicians with above-median ratings.[18] For physicians with 50% of pay rising in proportion to *RVUs* and with above-median ratings, this roughly 24% increase in *RVUs* for the median physician would increase pay by roughly 12%.

5.3 RATING IMPROVEMENT

The analysis from subsection 5.2 documents that rating disclosure yields a payoff for physicians who had higher ratings. Model 3, specified as follows, tests for related effects on performance by ratings.

$$Rating_{pytv} = \alpha + \delta Physician_p + \lambda Year_y + \omega Period_t + \varsigma Controls_v$$
$$+\beta Disclosed_{pt} + \varepsilon_{pytv}. \tag{3}$$

The model's subscripts are the same as those in model 1, except that I replace w with v to index individual patient visits and I use t to index periods segmenting disclosure events that begin with the disclosure's announcement. *Disclosed* is an indicator equal to 1 in the time period following the

---

[15] In each of my models, there are at least 78 clusters. See Bertrand, Duflo, and Mullainathan [2004] and Angrist and Lavy [2009] for references suggesting that this number of clusters is sufficient.

[16] As a robustness test, I use block-bootstrapped standard errors. This involves taking a random subsample and estimating the model repeatedly to arrive at standard errors. When a physician is omitted from a given random subsample, the model cannot be estimated because it requires a fixed effect for each physician. To successfully apply this method, I replace the fixed effect for these less-frequently observed physicians with an indicator variable for whether they were assigned to treatment. The results of the hypothesis tests remain statistically significant at similar levels, and are sometimes significant at stronger levels.

[17] My results for *RVUs Per Week* are robust to not winsorizing and also to winsorizing at the 95th or 99th percentiles.

[18] The estimated effect on visits and *RVUs* is not statistically significant for physicians with median or below-median ratings. Administrators suggested that an increase in the number of patients across the system could explain why the increase in visit volume for high-rated physicians does not require a decrease in visit volume for low-rated physicians.

TABLE 2

Effect of Physician-Rating Disclosure on Visits Per Week and RVUs Per Week

	(1)	(2)	(3)	(4)
	Visits Per Week		*RVUs Per Week*	
	High-Rated Doc	Med–Low-Rated Doc	High-Rated Doc	Med–Low-Rated Doc
Disclosed	2.40**	1.63	11.54***	2.24
	[2.54]	[1.38]	[3.90]	[0.39]
Gender	2.84*	0.17	5.00	−0.79
	[1.65]	[0.15]	[0.86]	[−0.16]
CCI	3.26***	3.12***	8.58***	8.28**
	[5.37]	[4.92]	[2.94]	[2.11]
Medicare or Medicaid	−1.43	1.87	4.68	17.05**
	[−0.78]	[1.44]	[0.56]	[2.43]
Age dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes
Physician dummies	Yes	Yes	Yes	Yes

This table presents estimates of the effect of physician-rating disclosure on the number of visits and *RVUs* in the given week (*Visits Per Week* and *RVUs Per Week*) for physicians with a high rating (an above-median consensus rating at the start of disclosure) and physicians with a median or low rating (a median or below-median consensus rating at the start of disclosure). *Age* dummies are as outlined by Newman and Newman [2014] to capture differences in psychological function associated with age. Standard errors are clustered at the specialty level. T-statistics are reported in brackets. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. *Visits Per Week* and *RVUs Per Week* (High-Rated Doc) *N* = 36,937, *Visits Per Week* and *RVUs Per Week* (Med–Low-Rated Doc) *N* = 34,480.

announcement of disclosure if the given physician was eventually included in disclosure during my sample. The other right-hand variables in model 3, other than those contained in the controls vector, are also the same as those in model 1. The controls vector consists of the charges for the visit, *RVUs, CCI,* the physician's visit count that week, whether the visit's insurer was Medicare or Medicaid, whether the visit was the patient's first to the physician, whether the patient speaks English, and patient gender and age.

To orient my models to estimate changes in physician behavior, rather than changes in physician composition, my sample selection requires that a physician was present both before and after the disclosure's announcement. Also, in estimating performance effects, I take into account that physicians who were not included in the disclosure, but who were close to the 30-rating threshold in the predisclosure period, would likely have expected that there was some possibility that they would have their ratings disclosed. These physicians may have acted, to some degree, as if they were treated despite being in the "control" group in my DiD models. To avoid categorizing these physicians as a control when they may have acted as though they were treated, all of my tests of performance effects exclude physicians who came within three ratings of the 30-rating threshold in the preperiod. My results are robust to including these physicians.

Table 3 displays the results of estimating model 3. Column 1 shows that *Rating* time trends are positive during my sample.[19] The coefficient on *Disclosed* in columns 2 and 3 shows an estimated positive and statistically significant effect of the disclosure on *Rating*. The results support H1—that disclosing physician ratings leads to improvement by ratings. In figure 3, I show estimates of the treatment effect in event time. Using all the controls in model 3 and with *Rating* as the dependent variable, the plotted estimates are coefficients on interaction terms between an indicator for the given time period and an indicator for whether the physician was included in disclosure at any point during my sample. The estimates are thus counterfactual treatment effect estimates in the predisclosure period and true treatment effect estimates in the postdisclosure period.[20]

The estimated effect on *Rating* is strongest in the six months following the disclosure's announcement. This is consistent with the idea that physicians anticipate herding and attempt to set a high pattern for later ratings to follow. The climb in ratings that occurs within a matter of months after the disclosure's announcement is plausible in light of the predicted mechanisms. Many physicians hold weekly meetings with their staff

---

[19] There are positive time trends in ratings. This allows estimating improvement in *Rating* that occurs in spite of, rather than due to, herding. Specifically, herding would weight a physician's rating toward the physician's past rating average, and so work against, rather than help to explain, an incremental rise in *Rating* in actuality and as measured with a DiD model.

[20] In these figures, I cluster standard errors at the specialty-time-period level, so that the effect estimate in each period adjusts for clustering within the specialty during the relevant time period.

TABLE 3
Effect of Physician-Rating Disclosure on Rating

	(1)	(2)	(3)
		Rating	
Disclosed		0.162***	0.155***
		[4.05]	[3.81]
Gender			−0.016***
			[−2.73]
Charges			−0.000
			[−0.25]
RVUs			0.002***
			[3.11]
CCI			0.033***
			[3.32]
Medicare or Medicaid			0.006
			[0.76]
Visits Per Week			0.000
			[0.57]
First Visit			−0.040***
			[−5.71]
English Speaking			0.091***
			[6.47]
Year dummies			
2012	0.026***	0.025***	0.020**
	[2.86]	[3.33]	[2.60]
2013	0.072***	0.025	0.019
	[6.95]	[1.39]	[1.05]
2014	0.088***	0.040**	0.035**
	[7.34]	[2.41]	[2.09]
Age dummies	No	Yes	Yes
Period dummies	No	Yes	Yes
Physician dummies	No	Yes	Yes

This table presents estimates of the effect of physician-rating disclosure on physician ratings (*Rating*). Column 1 presents isolated time trends and columns 2 and 3 vary the controls included. *Age* dummies are as outlined by Newman and Newman [2014] to capture differences in psychological function associated with age. Standard errors are clustered at the specialty level. T-statistics are reported in brackets. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. $N = 99,774$.

members. Some physicians described using these meetings shortly after the disclosure's announcement to take steps that could affect ratings from that point forward. Such steps included delegating clerical work to increase the time that a physician could spend with each patient, offering more chances for patients to ask questions at the end of visits, and assigning nurses to do "rounds" in the waiting room to provide updates on wait time.

I estimate that rating disclosure led to a roughly 0.15 increase in *Rating*. This is enough to lift a physician from the 50th percentile of predisclosure ratings to the 91st percentile of predisclosure ratings. That effect magnitude is economically significant in the sense that UUHC administrators often describe the system's performance by ratings in relative terms. For example, in media outlets and on the health care system's Web page (Lee
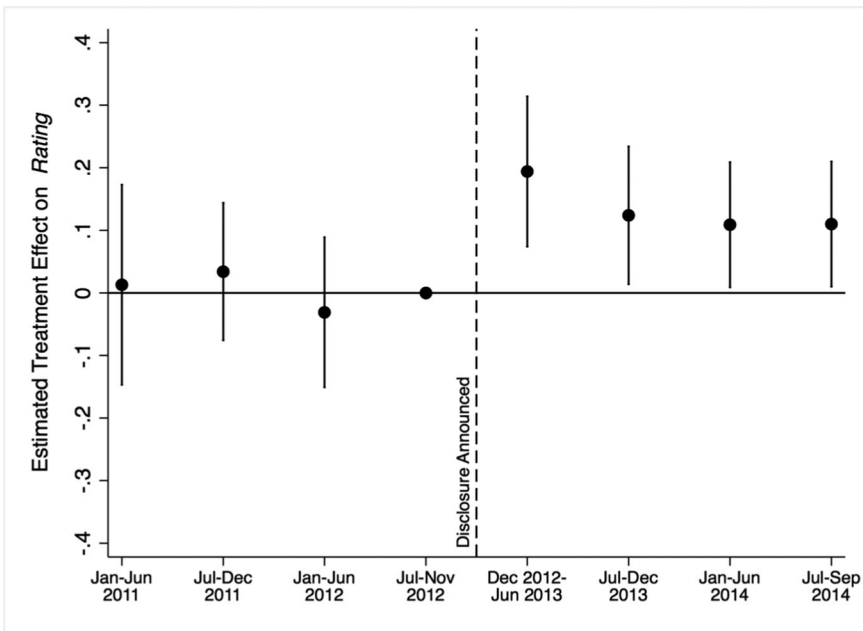
FIG. 3.—*Rating* effect estimated over time. This figure plots estimates of the effect of physician-rating disclosure on *Rating*, using indicator variables for time periods and their interactions with an indicator variable taking a value of 1 if the physician belonged to the group of physicians included in disclosure at some point during my sample. Dots represent coefficients on those interaction terms and vertical lines through dots represent 90% confidence intervals. The resulting plot shows the counterfactual treatment effect estimate in the predisclosure period and the true treatment effect estimate in the postdisclosure period.

[2014]), UUHC has highlighted its rise from the bottom 25th percentile of patient satisfaction ratings nationally in 2008 to above the 80th percentile nationally by the end of my sample period.

5.4 QUALITY IMPROVEMENT

Model 4, specified as follows, measures the effect of rating disclosure on the occurrence of objectively measured quality deductions including deviations from process or safety standards, readmissions, and hospital-acquired conditions.

$$Quality\ Deductions_{pytv} = \alpha + \delta Physician_p + \lambda Year_y + \omega Period_t + \varsigma Controls_v$$
$$+ \beta Disclosed_{pt} + \varepsilon_{pytv}. \tag{4}$$

The model's subscripts and the right-hand variables, other than those contained in the controls vector, are the same as those in model 3. The controls vector consists of the charges for the visit, *RVUs, CCI,* the

TABLE 4
Effect of Physician-Rating Disclosure on Quality Deductions

	(1)	(2)
	Quality Deductions	
Disclosed	−0.016**	−0.015**
	[−2.20]	[−2.19]
Gender		0.005*
		[1.94]
Charges		−0.000*
		[−1.87]
RVUs		−0.000
		[−0.03]
CCI		−0.006***
		[−3.26]
Medicare or Medicaid		0.005
		[1.07]
Visits Per Week		−0.000
		[−1.31]
Age dummies	Yes	Yes
Year dummies	Yes	Yes
Period dummies	Yes	Yes
Physician dummies	Yes	Yes

This table presents estimates of the effect of physician-rating disclosure on the sum of objectively measured quality deductions for a visit (*Quality Deductions*). Columns 1 and 2 vary the controls included. *Age* dummies are as outlined by IEA [2019] for research on physical health. Standard errors are clustered at the specialty level. T-statistics are reported in brackets.*, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. $N = 40,997$.

physician's visit count that week, whether the insurer was Medicare or Medicaid, and patient gender and age.

Table 4 displays the results of the test of model 4. The coefficient on *Disclosed* in columns 1 and 2 shows an estimated negative and statistically significant effect of the disclosure on *Quality Deductions*. The results support H2—that disclosing physician ratings positively affects the quality of care that the physician provides. Figure 4 shows the effect estimate mapped over time, using the same method as described for figure 3 but applied to *Quality Deductions*. As with *Rating*, the estimated effects on *Quality Deductions* are strongest shortly after the disclosure's announcement, which is consistent with the idea that the *Rating* and *Quality Deductions* effects occur through a common mechanism such as physician time spent with patients. Health economics studies have found that small increases in physician time with patients, such as an additional few minutes in an office visit or a brief follow-up phone call, can reduce adverse events, including those that I pick up in *Quality Deductions,* by roughly 20% (Harrison et al. [2011], Neprash [2016]). Given that increased physician time with patients is only one of the possible mechanisms for my results, it is plausible that my estimated effects would be in this range or greater.

In line with this reasoning, the estimated decrease in *Quality Deductions* is roughly 37%. To gauge the economic impact of this improvement, I focus
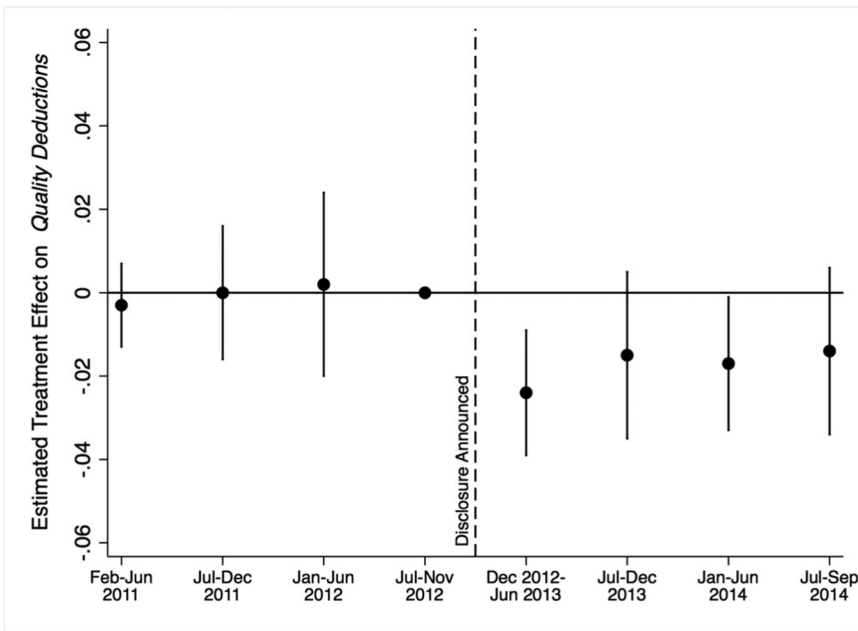
Fig. 4.—*Quality Deductions* effect estimated over time. This figure plots estimates of the effect of physician-rating disclosure on *Quality Deductions*, using indicator variables for time periods and their interactions with an indicator variable taking a value of 1 if the physician belonged to the group of physicians included in disclosure at some point during my sample. Dots represent coefficients on those interaction terms and vertical lines through dots represent 90% confidence intervals. The resulting plot shows the counterfactual treatment effect estimate in the predisclosure period and the true treatment effect estimate in the postdisclosure period.

on readmissions and hospital-acquired conditions and draw from research on the economic impact of these events. Analysis from the U.S. Department of Health and Human Services suggests that the average hospital-acquired condition leads to an increase in the chance of death by 7% and a cost of $22,257 per patient, and that a 30-day readmission leads to a cost of roughly $14,000 per patient (Healthcare Cost and Utilization Project [2015], DHHS [2017]).[21] Based on these per-patient estimates of economic effects of hospital-acquired conditions and 30-day readmissions, I estimate that the reduction in these occurrences in my setting saved over $20 million per year, and that the reduction in hospital-acquired conditions saved about 10 lives per year.[22]

---

[21] I report the average cost of a hospital-acquired condition weighted by the frequency with which each hospital-acquired condition type appears in my setting.

[22] A recent survey suggests that decreases in readmission rates can also reduce the instance of patient mortality (O'Malley, Alper, and Greenwald [2018]), but I cannot find a systematic survey of effect estimates as the Department of Health and Human Services has provided for hospital-acquired conditions (DHHS [2017]).

TABLE 5
Effect of Physician-Rating Disclosure on Time With Patient

	(1)	(2)
	Time With Patient	
Disclosed	4.38**	4.31***
	[2.59]	[2.63]
Gender		0.30***
		[2.80]
Charges		0.001***
		[3.42]
RVUs		−0.005
		[−0.44]
CCI		1.06***
		[4.83]
Medicare or Medicaid		−0.196
		[−1.40]
Visits Per Week		−0.003
		[−0.49]
Age dummies	Yes	Yes
Year dummies	Yes	Yes
Period dummies	Yes	Yes
Clinic dummies	Yes	Yes
Physician dummies	Yes	Yes

This table presents estimates of the effect of physician-rating disclosure on the amount of time in minutes that a physician spends interacting with a patient during a given patient visit (*Time With Patient*). Columns 1 and 2 vary the controls included. *Age* dummies are as outlined by IEA [2019] for research on physical health. Standard errors are clustered at the specialty level. T-statistics are reported in brackets. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. $N = 130{,}436$.

5.5 TIME WITH PATIENT

To understand a possible mechanism for the effects on *Rating* and *Quality Deductions*, I consider the amount of time that physicians spent with patients. I first demonstrate that rating disclosure affected the number of minutes that a physician spends with a patient on a given visit, or *Time With Patient*, prior to assessing whether such an effect mediates the effect of disclosure on *Rating* and *Quality Deductions*. To do this first step, I estimate model 4, replacing the dependent variable with *Time With Patient* and including clinic fixed effects. The coefficient on *Disclosed* in column 2 of table 5 suggests that the disclosure leads physicians to spend roughly an additional 4.31 minutes with each patient on average. Figure 5 shows the effect estimate mapped over time, using the same method that I applied to produce figures 3 and 4. Consistent with figures 3 and 4, the performance-effect estimates are strongest shortly after the disclosure's announcement. As discussed in subsection 5.3, these improvements would plausibly occur within six months, given that physicians reported responding to the disclosure's announcement by quickly taking steps to allocate more of their time to patients.

I then apply mediation analysis, as outlined by Sobel [1982] and Tofighi and MacKinnon [2016], to identify whether an increase in *Time With Patient*
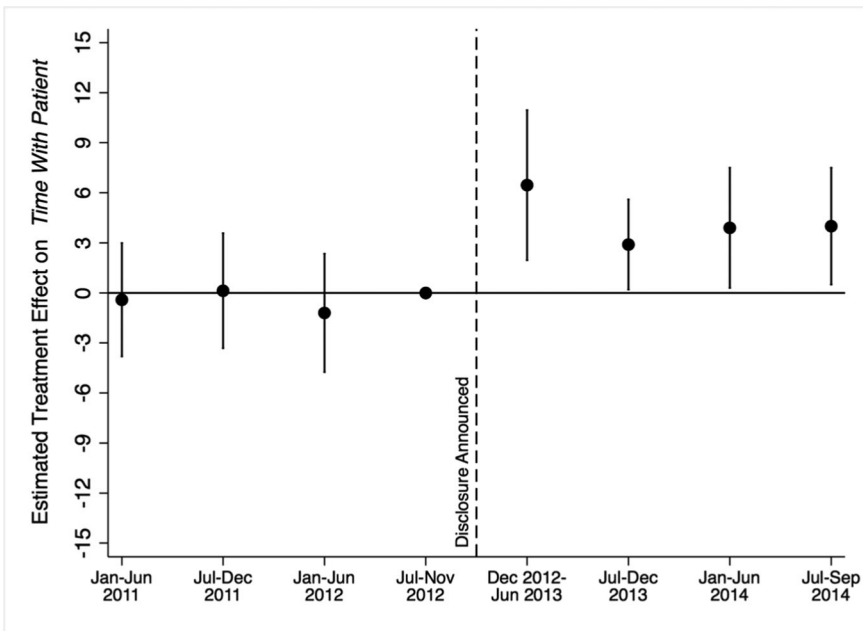
FIG. 5.— *Time With Patient* effect estimated over time. This figure plots estimates of the effect of physician-rating disclosure on *Time With Patient*, using indicator variables for time periods and their interactions with an indicator variable taking a value of 1 if the physician belonged to the group of physicians included in disclosure at some point during my sample. Dots represent coefficients on those interaction terms and vertical lines through dots represent 90% confidence intervals. The resulting plot shows the counterfactual treatment effect estimate in the predisclosure period and the true treatment effect estimate in the postdisclosure period.

mediates the estimated effects of rating disclosure on *Rating* and on *Quality Deductions*. Table 6 displays the results of estimating the portions of the total effects on *Rating* and *Quality Deductions* that operate when *Time With Patient* is specified as a mediating variable. This is the portion estimated to occur only when *Time With Patient* is allowed to vary as a result of rating disclosure. In the first row, I show the estimated total effect of rating disclosure on *Rating* and *Quality Deductions*, from models 3 and 4, respectively. Below those estimated total effects, and consistent with H3a (H3b), the mediation analysis estimates that roughly 16% (20%) of rating disclosure's total effect on *Rating* (*Quality Deductions*) operates via an increase in *Time With Patient*.

5.6 HERDING

I use two independent methodologies to estimate herding. The first methodology estimates herding as a tightening of a physician's rating distribution around his or her prior average rating when the prior average rating is disclosed (Eryarsoy and Piramuthu [2014]). *Absolute Difference* is the distance between a physician rating and the physician's prior average rating as calculated for disclosure. In the postdisclosure period, this prior

	(1)	(2)
	Rating	*Quality Deductions*
Estimated Effect of Physician-Rating Disclosure:		
Total effect	0.155***	−0.015**
	[3.81]	[−2.19]
Effect mediated by *Time With Patient*	0.025**	−0.003*
	[2.08]	[−1.87]
Percentage of total effect mediated by *Time With Patient*	16.71%	20.02%

This table presents estimates of the portion of the effect of physician-rating disclosure on *Rating* and *Quality Deductions* that is mediated by *Time With Patient*. The mediated portion is the portion that is estimated to operate through changes in *Time With Patient* (Sobel [1982]). Standard errors are clustered at the specialty level. T-statistics are reported in brackets. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

average rating is the physician's prior 12-month average as it stood at the most recent of December 2012 and July 2013. I calculate and update this prior average rating for the physician in the same manner and on the same schedule of December and July in the predisclosure period. I use model 5 to test for a reduction in *Absolute Difference* that occurs for disclosed physicians and not for undisclosed physicians, as follows:

$$Absolute\ Difference_{pytv} = \alpha + \delta Physician_p + \lambda Year_y + \omega Period_t$$
$$+ \varsigma Controls_v + \beta Disclosed_{pt} + \varepsilon_{pytv}. \qquad (5)$$

The model's subscripts are the same as those in model 1, except for the replacement of w with v to index individual patient visits. *Disclosed* is an indicator equal to 1 in the time period following which the given physician's ratings were disclosed, if ever during my sample. The controls vector includes all controls used in model 3, along with a control for the standard deviation of the physician's ratings in the given period relative to disclosure. This control exploits the fact that I can measure herding as a decreased distance around the particular average rating as it stood for a physician at the time of disclosure. By controlling for an overall tightening of the distribution of a physician's ratings after disclosure, and measuring the tightening of a physician's ratings around that particular average rating for the physician, my models help to rule out that the result is driven by factors such as a physician's service becoming more consistent.

To estimate model 5, I narrow the sample to the two-year range centered at the first disclosure. The starting point of this range is the earliest date at which I have 12 months of prior data, and is thereby the earliest date at which I can calculate a given physician's 12-month average rating as UUHC calculated it for disclosure. The ending point of this range is 12 months

TABLE 7
Effect of Physician-Rating Disclosure on Absolute Difference

	(1)	(2)
	Absolute Difference	
Disclosed	−0.024***	−0.029***
	[−2.62]	[−3.73]
Gender		0.007
		[1.64]
Charges		0.000
		[1.16]
RVUs		−0.001***
		[−4.24]
CCI		−0.013*
		[−1.69]
Medicare or Medicaid		0.003
		[0.81]
First Visit		0.021***
		[3.84]
Visits Per Week		0.000
		[0.12]
English Speaking		−0.071***
		[−4.15]
Standard Deviation	0.364***	0.361***
	[16.83]	[16.69]
Age dummies	Yes	Yes
Year dummies	Yes	Yes
Period dummies	Yes	Yes
Physician dummies	Yes	Yes

This table presents estimates of the effect of physician-rating disclosure on the absolute difference (*Absolute Difference*) of physician ratings from the given physician's prior consensus rating. Columns 1 and 2 vary the controls included. *Age* dummies are as outlined by Newman and Newman [2014] to capture differences in psychological function associated with age. T-statistics are reported in brackets. Standard errors are clustered at the specialty level. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. $N = 59{,}043$.

after the first disclosure, which ensures that the results are not attributable to measuring *Absolute Difference* over different lengths of time in the predisclosure and postdisclosure periods. I restrict the treatment group to physicians who qualified for the first date of disclosure, in December 2012, and exclude the small fraction who first qualified for disclosure on the second date of disclosure, in July 2013. This provides a 12-month period for measuring the formation of a herd around disclosed ratings in the postdisclosure period within the narrowed sample.

Table 7 displays the results of the test of model 5. The coefficient on *Disclosed* in columns 1 and 2 shows an estimated negative and statistically significant effect of disclosure on *Absolute Difference*. This supports H4a—that rating disclosure leads to herding as measured by a tightening of the rating distribution around a prior disclosed average rating. Figure 6 shows the effect estimate over time, using the same method that I used to produce figures 3–5.
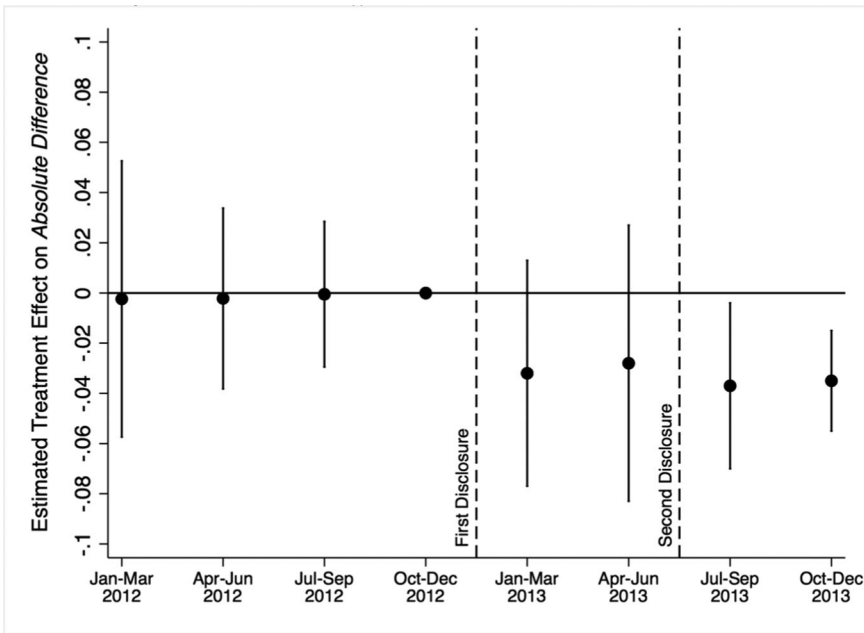
Fig. 6.—*Absolute Difference* effect estimated over time. This figure plots estimates of the effect of physician-rating disclosure on *Absolute Difference*, using indicator variables for time periods and their interactions with an indicator variable taking a value of 1 if the physician belonged to the group of physicians included in the first disclosure. Dots represent coefficients on those interaction terms and vertical lines through dots represent 90% confidence intervals. The resulting plot shows the counterfactual treatment effect estimate in the predisclosure period and the true treatment effect estimate in the postdisclosure period.

In a second, independent test of herding, I use data on objective measures. The first row of table 8 (table 9) shows that *Rating* rises with a decrease in *Quality Deductions* (increase in *Time With Patient*) in the predisclosure period.[23] This analysis controls for covariates in models 3 and 4. The relationships become statistically significantly weaker in the postdisclosure period, which is consistent with *Rating* becoming jammed near a past consensus and so becoming less responsive to changes in objective performance improvement. These results support H4b—that rating disclosure

---

[23] To measure the relationship between *Quality Deductions* and *Rating*, I estimate a physician's fixed effect for *Quality Deductions* and a physician's fixed effect for *Rating* in the given period relative to the start of disclosure, and then regress the physician fixed effects for *Rating* on the physician fixed effects for *Quality Deductions* in the given period. I describe this analysis in more detail below (table 8). This approach accounts for the relationship between *Quality Deductions* from visits that the physician conducts and *Rating* at later visits with the physician when patients may become aware of the impact of *Quality Deductions*. This approach also accounts for the relationship between *Rating* at visits that the physician conducts and *Quality Deductions* at later visits with the physician that the earlier rated visits were intended to serve as preparation for.

TABLE 8

Actual Ratings and Ratings Estimated Based on Reduced Quality Deductions

	Predisclosure	Postdisclosure	Predisclosure – Postdisclosure
β_{QD} :	−2.60***	−0.43	−2.17**
	(0.68)	(0.67)	[−2.27]

	Actual	Estimate	Estimated Effect of Herding on *Rating*
Postdisclosure	4.74***	4.77***	−0.035*
Rating:	(0.007)	(0.018)	[−1.75]

This table presents an estimate of herding as the difference between actual *Rating* and *Rating* estimated based on reduced *Quality Deductions*. The first row shows the relationship between *Quality Deductions* and *Rating*, denoted β_{QD}. I estimate a physician's fixed effect for *Quality Deductions* and a physician's fixed effect for *Rating* in the given period relative to the start of disclosure, and then measure β_{QD} by regressing the physician fixed effects for *Rating* on the physician fixed effects for *Quality Deductions* in the given period. The physician fixed effects for *Rating* (*Quality Deductions*) account for all of the controls in model 3 (model 4). This approach accounts for the relationship between *Quality Deductions* from visits that the physician conducts and *Rating* at later visits with the physician when patients may become aware of the impact of *Quality Deductions*. This approach also accounts for the relationship between *Rating* at visits that the physician conducts and *Quality Deductions* at later visits with the physician that the earlier rated visits were intended to serve as preparation for. To provide a sufficient sample size for measuring the physician's fixed effect, I restrict the sample for this analysis based on the physician's number of ratings received and procedures performed. I include only physicians who received more than 30 ratings in both the predisclosure and postdisclosure periods. Based on formulas for sample size in analysis of rare events with the incidence of *Quality Deductions* (Machin et al. [1997]), I include only physicians who performed more than 98 (85) procedures in the predisclosure (postdisclosure) period. The relationship between *Rating* and *Quality Deductions* has a smaller magnitude in the postdisclosure period—that is, *Rating* rises less with a decrease in *Quality Deductions* after physician-rating disclosure. Row 2 shows the average actual *Rating* for physicians included in disclosure and then an estimate of what that average *Rating* would have been given the reduction in *Quality Deductions* and the predisclosure β_{QD} that is not affected by herding. The difference between actual *Rating* and the estimate of *Rating* provides an estimate of the degree to which herding keeps *Rating* suppressed at prior levels even as underlying quality improves. Standard errors are clustered at the specialty level. T-statistics are reported in brackets and standard errors are reported in parentheses. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

leads to herding as measured by a decrease in the responsiveness of ratings to changes in objective measures of underlying quality.

Row 2 of table 8 (table 9) shows that if the predisclosure relationship between *Rating* and *Quality Deductions* (*Time With Patient*) had held in the postdisclosure period, then, given the estimated effect of the disclosure on *Quality Deductions* (*Time With Patient*), *Rating* for disclosed physicians would have risen to an average of 4.77 after disclosure. Instead, *Rating* for these physicians rises only to an average of 4.74. In figure 7 (figure 8), I map the difference between actual ratings for disclosed physicians and what the ratings would have been had the predisclosure relationship between *Quality Deductions* (*Time With Patient*) held in the postdisclosure period. The results in tables 8 and 9 suggest that herding keeps ratings about 0.035 points closer to a physician's posted consensus after it is disclosed. This estimate of herding is of similar magnitude to the estimated 0.029 magnitude from table 7. In light of the 0.15 estimated effect of disclosure on ratings, my herding estimates of around 0.029 to 0.035 suggest that herding dampens the effect of rating disclosure on rating improvement by roughly 20%.

TABLE 9

Actual Ratings and Ratings Estimated Based on Increased Time With Patient

	Predisclosure	Postdisclosure	Predisclosure – Postdisclosure
β_{TWP} :	0.010***	0.002**	0.008**
	(0.003)	(0.001)	[2.52]

	Actual	Estimate	Estimated Effect of Herding on *Rating*
Postdisclosure	4.74***	4.77***	−0.034*
Rating:	(0.007)	(0.021)	[−1.78]

This table presents an estimate of herding as the difference between actual *Rating* and *Rating* estimated based on increased *Time With Patient*. The first row shows the relationship between *Rating* and *Time With Patient*, denoted β_{TWP}. This is the coefficient on *Time With Patient* from a regression of *Rating* on *Time With Patient* using all of the controls common to models 3 and 4, with *Age* coded as outlined in Newman and Newman [2014]. The results for β_{TWP} are statistically significant at the same levels and equivalent to three decimal places when I use *Age* coded as outlined in IEA [2019]. The relationship between *Rating* and *Time With Patient* has a smaller magnitude in the postdisclosure period—that is, *Rating* rises less with an increase in *Time With Patient* after physician-rating disclosure. Row 2 shows the average actual *Rating* for physicians included in disclosure and then an estimate of what that average *Rating* would have been given the increase in *Time With Patient* and the predisclosure β_{TWP} that is not affected by herding. The difference between actual *Rating* and the estimate of *Rating* provides an estimate of the degree to which herding keeps *Rating* suppressed at prior levels even as *Time With Patient* increases. Standard errors are clustered at the specialty level. T-statistics are reported in brackets and standard errors are reported in parentheses. *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table 10 displays the results of tests for whether herding in ratings leads to dynamic incentives. H5 predicted that physicians would anticipate herding and work to set a favorable rating pattern by concentrating on initial performance. The estimated performance improvement by *Quality Deductions* and *Time With Patient* is of larger magnitude during the first six months following disclosure's announcement than during the full postdisclosure period.

Below these estimates in table 10, I run the same tests in cross-sections of *Pageviews* and *Consensus Rating Count*. *Pageviews* are the number of pageviews of the physician's Web page in the calendar month prior to the observed visit. *Consensus Rating Count* is the sample size of the physician's consensus rating as UUHC calculated it for disclosure. As noted in section 2, physicians would plausibly anticipate that rating disclosure will lead to more pronounced herding when they have an above-median *Consensus Rating Count* or above-median predisclosure *Pageviews*.[24] Table 10 shows that, under these conditions, the shift in effort is more dramatic. Specifically, the estimated effects of disclosure on *Time With Patient* and *Quality Deductions* have relatively larger magnitudes in the first six months under these conditions, and these magnitudes reach a level that is statistically significantly larger than in the last six months of the sample.

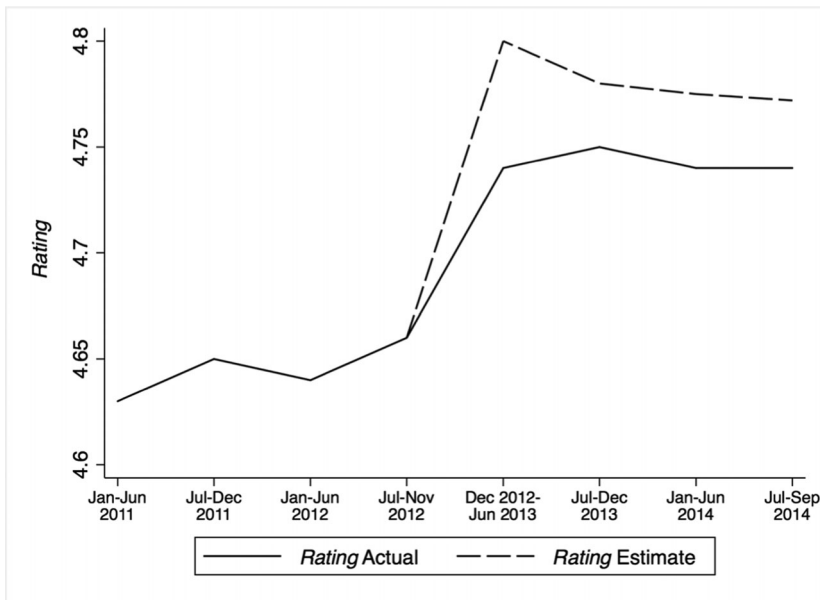[24] In untabulated analyses, I find evidence that herding is more pronounced under these conditions.

FIG. 7.—*Rating* actual and as estimated based on *Quality Deductions*. This figure plots the actual average *Rating* for physicians included in disclosure and an estimate of what that average *Rating* would have been if there were no herding in the postdisclosure period. To construct this estimate, I first quantify the relationship between *Rating* and *Quality Deductions* in the predisclosure period and then use that relationship to predict *Rating* in the postdisclosure period based on changes in *Quality Deductions*. Actual ratings stay closer to their prior values than do ratings as estimated based on the improvement in quality, consistent with herding among raters. The difference between the *Rating* Estimate and *Rating* Actual lines provides an estimate of herding, which is tabulated in table 8.

5.7 ROBUSTNESS TESTS

A key identifying assumption in the DiD models that I apply is that dependent variables would have trended similarly, absent disclosure, for physicians who were included in and those who were excluded from disclosure. Tests of parallel trends are inherently limited to the pretreatment period, given that it is not possible to observe the treatment group in the posttreatment period absent treatment.

I have two years of data prior to disclosure, which provides a relatively short time for establishing parallel trends in quality if changes in quality occur annually. Figures 3–5 document that there is no divergence in trends between years one and two in the pretreatment period, which helps to mitigate concern about annual changes, albeit subject to the assumption that these changes occur every year. I also break down the data within years in these figures to provide evidence that the trends stay flat within and across roughly two years prior to the disclosure and then shift relatively shortly after disclosure.
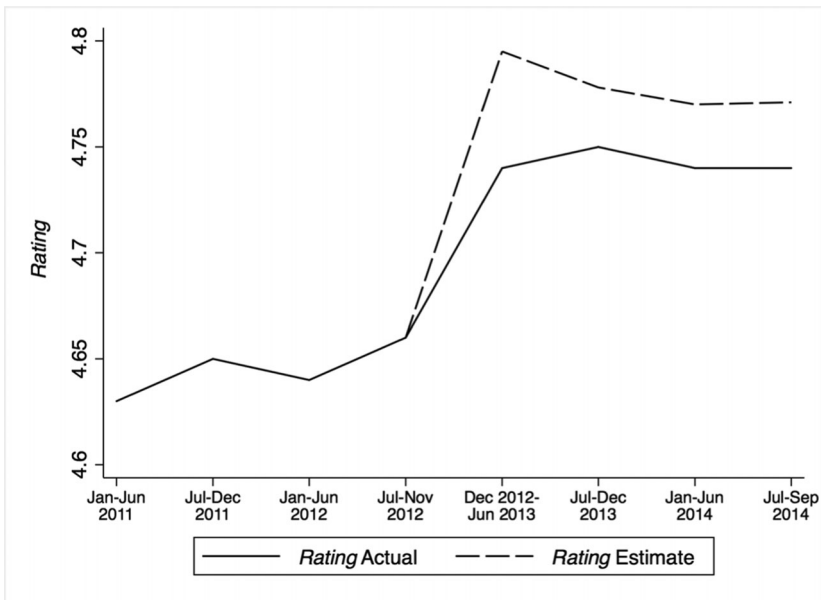
Fig. 8.—*Rating* actual and as estimated based on *Time With Patient*. This figure plots the actual average *Rating* for physicians included in disclosure and an estimate of what that average *Rating* would have been if there were no herding in the postdisclosure period. To construct this estimate, I first quantify the relationship between *Rating* and *Time With Patient* in the predisclosure period and then use that relationship to predict *Rating* in the postdisclosure period based on changes in *Time With Patient*. Actual ratings stay closer to their prior values than do ratings as estimated based on the increase in *Time With Patient*, consistent with herding among raters. The difference between the *Rating* Estimate and *Rating* Actual lines provides an estimate of herding, which is tabulated in table 9.

Additionally, I conduct tests of placebo disclosure prior to actual disclosure. I assign treatment based on whether a physician had received the necessary 30 ratings to qualify for the placebo disclosure. I show the results of this test, along with the results of my other robustness tests, in the online appendix. Table A1 of the online appendix displays null results in the predisclosure period when I use these placebo-disclosure dates.

As a second placebo test, I randomly assign physicians to the groups included in and excluded from disclosure and then estimate the effects of disclosure. Fewer than 5% of the estimated coefficients on *Disclosed* are statistically significant at the 0.05 level when I repeat this process 1,000 times.

My models control for several observables. These include the number of years that a physician has been with UUHC (*Years With UUHC*), which administrators noted would likely be fewer for Second-Disclosed Physicians. Table 1 shows that physicians in the Second-Disclosed group are a small number who are much newer to UUHC than the First-Disclosed or Nondisclosed groups. When I rerun my analyses excluding these physicians, the results remain statistically significant and have similar magnitudes.

TABLE 10

Effect of Physician-Rating Disclosure on Early Performance by Objective Measures

	Time With Patient	*Quality Deductions*
Estimated Effect of Physician-Rating Disclosure Six Months After Disclosure Announcement:		
Full sample	6.45**	−0.023**
	[2.41]	[−2.12]
Above-median *Consensus Rating Count*	9.02**,††	−.033**,††
	[2.34]	[−2.13]
Above-median *Pageviews*	8.76**,††	−.028**,††
	[2.22]	[−2.35]

This table presents estimates of the effect of physician-rating disclosure on *Time With Patient* and *Quality Deductions* for the first six months following the disclosure's announcement, in the full sample and then in the sample of physicians who had an above-median *Consensus Rating Count* and the sample of physicians who had above-median predisclosure *Pageviews*. The estimates are from the fully specified estimation of model 4, as described in subsections 5.4 and 5.5 for application to *Quality Deductions* and *Time With Patient*, respectively. Standard errors are clustered at the specialty level. T-statistics are reported in brackets. *, **, and *** denote coefficients' significance at the 0.1, 0.05, and 0.01 levels, respectively. †, ††, and ††† denote that the effect estimate is statistically significantly larger in the six months following the disclosure's announcement than in the last six months of the sample at the 0.1, 0.05, and 0.01 levels, respectively.

As an additional means of controlling for observables, I use propensity-score matching to construct matched samples of physicians included in and excluded from disclosure that are similar by observables including *Years With UUHC*. Table A2 of the online appendix shows the descriptive statistics of the matched samples. Table A3 of the online appendix shows that my results hold among these matched samples.

I also seek to address the concern that differences in the groups that I compare may be due to differences in medical specialties. Physicians from specialties that involve more visits are generally more likely to receive enough ratings to qualify for disclosure. As a robustness test, I randomly match each physician excluded from disclosure with a physician from the same specialty who was included in disclosure. Table A4 of the online appendix shows that my results hold when there are equal numbers of physicians from each specialty in the groups excluded from and included in disclosure.

Although physician fixed effects preclude time-invariant differences from affecting my estimates, omitted variables that are correlated with a physician qualifying for disclosure may affect my estimates. To help address this concern, I restrict my sample to physicians within the band of 20–40 received ratings in the prior 12 months at the start of disclosure. Relative to physicians from my full sample, these physicians should be more similar by unobservable variables correlated with the number of received ratings. Table A5 of the online appendix shows that my results hold with this restriction.

I also perform tests to address the concern that the disclosure's effect on *Rating* results from survey-response bias. The response rate during my sample is roughly 5% and does not change significantly after disclosure either

for physicians included in or excluded from disclosure. I partition the sample for *Rating* by whether a physician's response rate increased following disclosure. An increase in survey response rate could indicate that a physician is encouraging responses from certain patients. Table A6 of the online appendix shows that the improvement in *Rating* holds at similar levels when a physician's response rate did not increase.

I conduct an additional robustness test to control for effects of annual hiring on quality. Specifically, I exclude physicians who were hired during my sample. Table A7 of the online appendix shows that my results are robust to excluding these physicians. Collectively, my robustness tests offer support for the assumptions of my empirical design.

5.8 generalizability

A few distinctive features of health care are important to account for when considering the generalizability of my results to other settings. First, demand may be relatively less sensitive to disclosure in health care because frictions make it hard for patients to switch among physicians. Though patients can select among physicians at UUHC, they may default to the option suggested by a referring physician or an appointment scheduler. Second, physicians in my setting are employees and so do not have rights to all of the profit they produce. They are also protected by barriers to entry, including advanced training and licensure. This limits how much a physician can gain or lose financially as a result of changes in demand. If demand is more sensitive to consumer ratings in other settings and the rated entity has more to gain or lose financially, performance effects of the ratings' public visibility could be larger than those that I find.

Third, health care is subject to high levels of information asymmetry (Arrow [1963]). When individuals have less private information, they may be more likely to rely on a public rating to form their own rating. The limited private information in health care may make herding more likely than it would be in settings in which private information is more complete.

I am also limited to data from one organization. However, the results may be relatively generalizability to other organizations for a couple of reasons. First, the data span tens of thousands of patients in a health care system whose referral area spans five states. Second, the disclosure in my setting resembles disclosures from a variety of other organizations, whereby the organization gathers and then publicly discloses subjective ratings regarding its service professionals (e.g., Stanford Health Care [2016], Redfin [2018], Texas Tech University [2020]).

6. Conclusion

This study examines performance effects of the public disclosure of subjective ratings. Using data from a health care system that disclosed patient

ratings of its physicians, I demonstrate broad, positive performance effects. These effects include better performance by the ratings and by objective measures of medical quality. I also find evidence of a few drawbacks of rating disclosure. First, the ratings become jammed near initial values, a result that information cascade theory helps to explain. Second, physicians seem to anticipate rating jamming and respond by shifting effort toward earlier performance.

My study's main contribution is to the literature on disclosure's real effects, which has not previously examined effects of disclosing subjective ratings. This type of disclosure has spread to many industries and can significantly affect demand. I provide initial evidence of the effects of subjective rating disclosure on performance. My analysis also lends support to theories regarding multitasking, information cascade, and subjective performance evaluation.

APPENDIX A: PATIENT SATISFACTION SURVEY QUESTIONS USED IN DISCLOSURE

Physician (referred to as "care provider" in the survey questions)

1. Friendliness/courtesy of the care provider
2. Explanations the care provider gave you about your problem or condition
3. Concern the care provider showed for your questions or worries
4. Care provider's efforts to include you in decisions about your treatment
5. Degree to which care provider talked with you using words you could understand
6. Amount of time the care provider spent with you
7. Your confidence in this care provider
8. Likelihood of your recommending this care provider to others
9. Length of wait time at clinic

© 2016 Press Ganey Associates, Inc.

APPENDIX B: EXAMPLE PATIENT COMMENTS

Each patient satisfaction survey contains a text box under the Care Provider section of the survey with the prompt: "Comments (Describe Good or Bad Experience)" that patients can choose to fill in. Comments regarding physicians were posted in their entirety on the official online profiles of physicians included in disclosure, except when administrators filtered out comments that contained profanity, slander, or personally

identifiable information about the patient. The following are example comments:

Number	Selected Comment
1	The only complaint I had was that he did not tell me how many precancer spots I had on my face so I was not prepared for as many as there were. I was not quite mentally prepared for getting sprayed with the liquid nitrogen that many times.
2	Excellent service all the way around.
3	Dr. Salari is the best physician there in my opinion. I have not really seen any others but I trust him to give me an honest opinion and he shows he cares. Apparently everyone likes him because at times he is hard to get an appointment with but that I guess is a good thing.
4	I would have liked to have heard a bit about the down-side of "prednisone"—such as it can sometimes cause mood swings or depression. I was not prepared for "feeling so down."
5	The symptoms I had were frightening and the physician was very good at explaining what was going on and alleviating my fears.

APPENDIX C: VARIABLE DEFINITIONS

Dependent Variables	Description
Rating	The average of the nine component ratings regarding a physician in a Press Ganey survey returned following a patient visit. Each component rating is on a Likert scale of 1–5, with 5 as the most favorable rating.
Quality Deductions	The sum of the following quality deductions that the hospital measures and seeks to minimize: a deviation from a process or safety standard, a readmission to the emergency department within three months of discharge, or a hospital-acquired condition.
Absolute Difference	The absolute value of the difference between the rating that a physician receives for a visit and the physician's consensus rating. A physician's consensus rating, as UUHC calculated it for disclosure and as I use in my analysis, is the physician's prior 12-month average at the time of a disclosure. In the preperiod, I measure *Absolute Difference* from the consensus rating calculated on the same dates of the year on which disclosure occurred in the postperiod. I measure *Absolute Difference* in the same manner for physicians whether or not they were included in disclosure.
Time With Patient	The number of minutes that a physician spent with a patient during a visit as measured using timestamps from the patient's electronic health record.
Visits Per Week	The number of visits that a physician conducted in the given week.

Dependent Variables	Description
RVUs Per Week	The total number of *RVUs* (relative value units) that a physician generated from his or her visits in the given week. The number of *RVUs* for a visit reflects the severity of the case and the related complexity of treatment. The number is based on Medicare codes for the case, including diagnosis and patient demographics, and codes for related treatment. The allowable Medicare reimbursement for a visit rises proportionally to *RVUs*.

Treatment Variables	Description
Disclosed	In tests of performance effects, this is an indicator for the time period starting at the announcement of disclosure for physicians eventually included in disclosure. In tests of revenue, visit volume, and herding, this is an indicator for the time period starting when a given physician's ratings were disclosed, if ever during my sample.
Placebo Disclosed	In tests of performance effects, visit volume, and revenue, this is an indicator for the time period starting one year prior to the first disclosure for physicians who had received 30 ratings in the 12 months preceding this placebo disclosure. In tests of herding, this is an indicator for the time period starting six months prior to the first disclosure for physicians who had received 30 ratings in the 12 months preceding this placebo disclosure.

Partitioning Variables	Description
Consensus Rating Count	The count of observations that comprise a physician's consensus rating as calculated for disclosure and used in this study in measuring *Absolute Difference.*
Pageviews	The number of pageviews of the physician's official online profile in the calendar month prior to the observed visit.

Control Variables	Description
Age	Patient age at the time of the visit, with ages above 89 treated as 90. For the tests of models 1, 2, 3, and 5, ages are included as indicator variables using the psychometric categories of Newman and Newman [2014]: 0–11, 12–17, 18–24, 25–34, 35–59, 60–74, 75+. For the tests of model 4, ages are included as indicator variables using guidance from IEA [2019] for research on physical health: 0–4, 5–14, 15–24, 25–34, 35–44, 45–54, 55–59, 60+.

Control Variables	Description
Gender	An indicator variable equal to 1 if the patient is female.
Medicare or Medicaid	An indicator variable equal to 1 if Medicare or Medicaid was the primary insurance used for the visit.
RVUs	The number of *RVUs* (relative value units) for the visit. The number of *RVUs* for a visit reflects the severity of the case and the related complexity of treatment. The number is based on Medicare codes for the case, including diagnosis and patient demographics, and codes for related treatment. The allowable Medicare reimbursement for a visit rises proportionally to *RVUs*.
CCI	The Charlson Comorbidity Index (*CCI*), which takes a value of 1, 2, 3, or 6 in proportion to the likelihood of mortality within one year associated with the comorbid condition. Comorbid conditions include heart disease, diabetes, and cancer among the 22-condition set. The conditions are recorded at the time of a procedure. Analyses of objectively measured quality regard procedures and are thus able to include *CCI* as measured at the given visit. Ratings and time with patient are measured for visits regardless of whether the visit included a procedure, and thereby regardless of whether *CCI* is measured. For analyses of ratings and time with patient, I include *CCI* as its value for the patient in UUHC visits during the six-month window centered at the rated visit. The results are robust to narrowing this window to three months or expanding it to one year. For analyses of weekly *RVUs* or visits, I include *CCI* as measured on average for the physician during the given week.
Charges	The dollar value of charges assigned to the visit.
First Visit	An indicator variable equal to 1 if the visit is the patient's first to the physician conducting the visit.
Visit Per Week	The number of visits that a physician conducted in the given week.
English Speaking	An indicator variable equal to 1 if a patient indicated in a survey response that the patient speaks English.
Standard Deviation	The standard deviation of the physician's ratings in the period in which the rating occurred relative to disclosure.
Year	A categorical variable for the calendar year in which the visit occurred: 2011, 2012, 2013, or 2014.
Period	In tests of revenue, visit volume, and herding, this is a categorical variable for the period, segmented by disclosure events (i.e., the December 2012 and the July 2013 rating postings), in which the visit occurred: 1 for before the first posting, 2 for after the first and before the second posting, and 3 for after both postings. In tests of performance effects, this variable has the same coding except that segment 1 ends and segment 2 begins in November 2012, at the time of the disclosure's announcement.
Clinic	An indicator variable for the clinic at which the visit occurred.
Physician	An indicator variable for the physician conducting the visit.

Physician Variables	Description
Age	The physician's age as of January 1, 2015.
Gender	An indicator variable equal to 1 if the physician is female.
MD	An indicator variable equal to 1 if the physician holds an MD.
Years With UUHC	The number of years that UUHC has employed the physician.
Tenure Track	An indicator variable equal to 1 if the physician is on a tenure track at the University of Utah Medical School.

REFERENCES

ACEMOGLU, D.; T. A. HASSAN; AND A. TAHOUN. "The Power of the Street: Evidence from Egypt's Arab Spring." *The Review of Financial Studies* 31 (2018): 1–42.

ANDEL, C.; S. L. DAVIDOW; M. HOLLANDER; AND D. A. MORENO. "The Economics of Health Care Quality and Medical Errors." *Journal of Health Care Finance* 39 (2012): 1–11.

ANDERSON, L. R.; AND C. A. HOLT. "Information Cascades in the Laboratory." *The American Economic Review* 87 (1997): 847–62.

ANGRIST, J.; AND V. LAVY. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial." *The American Economic Review* 99 (2009): 1384–414.

ARROW, K. J. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53 (1963): 941–73.

BANERJEE, A. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107 (1992): 797–817.

BERTRAND, M.; E. DUFLO; AND S. MULLAINATHAN. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (2004): 249–75.

BEYER, A.; AND I. GUTTMAN. "Voluntary Disclosure, Manipulation, and Real Effects." *Journal of Accounting Research* 50 (2012): 1141–77.

BIKHCHANDANI, S.; D. HIRSHLEIFER; AND I. WELCH. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (1992): 992–1026.

BOL, J. C. "The Determinants and Performance Effects of Managers' Performance Rating Biases." *The Accounting Review* 86 (2011): 1549–75.

BOL, J. C.; T. M. KEUNE; E. M. MATSUMURA; AND J. Y. SHIN. "Supervisor Discretion in Target Setting: An Empirical Investigation." *The Accounting Review* 85 (2010): 1861–86.

BOND, P. "Panel discussion of 'The Real Effects of Measurement and Disclosure'." 50th Annual *Journal of Accounting Research* Conference. 2015. Available at https://research.chicagobooth.edu/arc/journal-of-accounting-research/jar-annual-conference/conference-web-casts#simple5

BOUWENS, J.; AND P. KROOS. "Target Ratcheting and Effort Reduction." *Journal of Accounting and Economics* 51 (2011): 171–85.

CASAS-ARCE, P.; AND F. A. MARTÍNEZ-JEREZ. "Relative Performance Compensation, Contests, and Dynamic Incentives." *Management Science* 55 (2009): 1306–20.

CANNIZZARO, A.; AND R. WEINER. "Does Dodd-Frank Disclosure Regulation Benefit Investors? Theory, Landscape, and Application to Extractive Industries." 2016. Available at https://www.sec.gov/comments/s7-25-15/s72515-22.pdf

CHANDRA, A.; J. GRUBER; AND R. MCKNIGHT. "Patient Cost-Sharing and Hospitalization Offsets in the Elderly." *The American Economic Review* 100 (2010): 193–213.

CHEVALIER, J. A.; AND D. MAYZLIN. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43 (2006): 345–54.

CHRISTENSEN, H. B.; E. FLOYD; L. Y. LIU; AND M. MAFFETT. "The Real Effects of Mandated Information on Social Responsibility in Financial Reports: Evidence from Mine-Safety Records." *Journal of Accounting and Economics* 64 (2017): 284–304.

CHRISTENSEN, H. B.; E. FLOYD; AND M. MAFFETT. "The Only Prescription is Transparency: The Effect of Charge-Price-Transparency Regulation on Healthcare Prices." *Management Science* 66 (2020): 2861–82.

CITY PERFORMANCE TEAM. "2019 City Survey." 2019. Available at https://sfgov.org/citysurvey/government-performance

DAFNY, L. S. "How Do Hospitals Respond to Price Changes?" *The American Economic Review* 95 (2005): 1525–47.

DANCIU, I.; J. D. COWAN; M. BASFORD; X. WANG; A. SAIP; S. OSGOOD; J. SHIREY-RICE; J. KIRBY; AND P. A. HARRIS. "Secondary Use of Clinical Data: The Vanderbilt Approach." *Journal of Biomedical Informatics* 52 (2014): 28–35.

DHHS. "Estimating the Additional Hospital Inpatient Cost and Mortality Associated with Selected Hospital-Acquired Conditions." 2017. Available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/haccost2017.html

DOYLE, J. J., JR. "Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far from Home." *American Economic Journal: Applied Economics* 3 (2011): 221–43.

DRANOVE, D.; D. KESSLER; M. McCLELLAN; AND M. SATTERTHWAITE. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *The Journal of Political Economy* 111 (2003): 555–88.

DUFLO, E. "Empirical Methods." Massachusetts Institute of Technology. 2002. Available at http://dspace.mit.edu/bitstream/handle/1721.1/49516/14-771Fall-2002/NR/rdonlyres/Economics/14-771Development-Economics-Microeconomic-Issues-and-Policy-ModelsFall2002/2494CA2C-D025-40A6-B167-F8A5662520DB/0/emp_handout.pdf

EBAY. "Seller Ratings." 2020. Available at https://www.ebay.com/help/buying/resolving-issues-sellers/seller-ratings?id=4023

ERYARSOY, E.; AND S. PIRAMUTHU. "Experimental Rating of Sequential Bias in Online Customer Reviews." *Information and Management* 51 (2014): 964–71.

EYSTER, E.; AND M. RABIN. "Naive Herding in Rich-information Settings." *American Economic Journal: Microeconomics* 2 (2010): 221–43.

FELTHAM, G.; AND J. XIE. "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations." *The Accounting Review* 69 (1994): 429–53.

FRIEDBERG, M. W.; D. G. SAFRAN; AND E. C. SCHNEIDER. "Satisfied to Death: A Spurious Result?" *Archives of Internal Medicine* 172 (2012): 1110–14.

GLOVER, L. "Are Online Physician Ratings any Good?" *US News*, 2014. Available at http://health.usnews.com/heath-news/patient-advice/articles/2014/12/19/are-online-physician-ratings-any-good

GRANJA, J. "Disclosure Regulation in the Commercial Banking Industry: Lessons from the National Banking Era." *Journal of Accounting Research* 56 (2018): 173–216.

HACHEM, F.; J. CANAR; F. FULLAM; A. S. GALLAN; AND S. HOHMANN. "The Relationships Between HCAHPS Communication and Discharge Satisfaction Items and Hospital Readmissions." *Patient Experience Journal* 1 (2014): 71–77.

HANAUER, D. A.; K. ZHENG; D. C. SINGER; A. GEBREMARIAM; AND M. M. DAVIS. "Public Awareness, Perception, and Use of Online Physician Rating Sites." *Journal of the American Medical Association* 311 (2014): 734–735.

HARRISON, P. L.; P. A. HARA; J. E. POPE; M. C. YOUNG; AND E. Y. RULA. "The Impact of Postdischarge Telephonic Follow-Up on Hospital Readmissions." *Population Health Management* 14 (2011): 27–32.

HEALTHCARE COST AND UTILIZATION PROJECT. "All-Cause Readmissions by Payer and Age, 2009–2013." 2015. Available at https://www.hcup-us.ahrq.gov/reports/statbriefs/sb199-Readmissions-Payer-Age.pdf

HEALY, P. M.; AND K. G. PALEPU. "Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature." *Journal of Accounting and Economics* 31 (2001): 405–40.

HOLMSTROM, B.; AND P. MILGROM. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7 (1991): 24–52.

HONG, F.; T. HOSSAIN; J. A. LIST; AND M. TANAKA. "Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories." National Bureau of Economic Research Paper No. w19660, 2013.

HRIBAR, M. R.; S. READ-BROWN; L. REZNICK; L. LOMBARDI; M. PARIKH; T. R. YACKEL; AND M. F. CHIANG. "Secondary Use of EHR Timestamp Data: Validation and Application for Workflow Optimization." *AMIA Annual Symposium Proceedings* 2015 (2015): 1909–17.

IEA. "Instructions to Authors." 2019. Available at https://academic.oup.com/ije/pages/Instructions_to_Authors

JIN, G. Z.; AND P. LESLIE. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *The Quarterly Journal of Economics* 118 (2003): 409–451.

JOYNT, K. E.; E. J. ORAV; AND A. K. JHA. "Thirty-day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." *Journal of the American Medical Association* 305 (2011): 675–81.

LEE, T. H. "Online Reviews Could Help Fix Medicine." *Harvard Business Review*. 2014. Available at https://hbr.org/2014/06/online-reviews-could-help-fix-medicine

LEE, Y.; K. HOSANAGAR; AND Y. TAN. "Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings." *Management Science* 61 (2015): 2241–58.

LEUZ, C.; AND P. WYSOCKI. "The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research." *Journal of Accounting Research* 54 (2016): 525–622.

LIN, C. T.; G. A. ALBERTSON; L. M. SCHILLING; E. M. CYRAN; S. N. ANDERSON; L. WARE; AND R. J. ANDERSON. "Is Patients' Perception of Time Spent with the Physician a Determinant of Ambulatory Patient Satisfaction?" *Archives of Internal Medicine* 161 (2001): 1437–42.

LINDBECK, A.; AND D. SNOWER. "Multitask Learning and the Reorganization of Work: From Tayloristic to Holistic Organization." *Journal of Labor Economics* 18 (2000): 353–76.

MACHIN, D.; M. CAMPBELL; P. FAYERS; AND A. PINOL. *Sample Size Table for Clinical Studies*, Second edition. Hoboken, NJ: Blackwell Science, 1997.

MUCHNIK, L.; S. ARAL; AND S. J. TAYLOR. "Social Influence Bias: A Randomized Experiment." *Science* 341 (2013): 647–51.

MULLEN, K. J.; R. G. FRANK; AND M. B. ROSENTHAL. "Can You Get What You Pay For? Pay-for-Performance and the Quality of Healthcare Providers." *The Rand Journal of Economics* 41 (2010): 64–91.

NEPRASH, H. T. "Better Late than Never? Physician Response to Schedule Disruptions." Working paper, 2016. Harvard University. Available at https://scholar.harvard.edu/files/hannahneprash/files/neprash_jmp_november2016.pdf

NEWMAN, B. M.; AND P. R. NEWMAN. *Development Through Life: A Psychosocial Approach*. Boston, MA: Cengage Learning, 2014.

O'MALLEY, T. A.; E. ALPER; AND J. GREENWALD. "Hospital Discharge and Readmission." UpToDate. 2018. Available at https://www.uptodate.com/contents/hospital-discharge-and-readmission

REDFIN. "Real-Estate Agents." 2018. Available at https://www.redfin.com/county/340/CA/San-Francisco-County/real-estate/agents

SCHLOSSER, A. E. "Posting Versus Lurking: Communicating in a Multiple Audience Context." *Journal of Consumer Research* 32 (2005): 260–65.

SINSKY, C.; L. COLLIGAN; L. LI; M. PRGOMET; S. REYNOLDS; L. GOEDERS; J. WESTBROOK; M. TUTTY; AND G. BLIKE. "Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties." *Annals of Internal Medicine* 165 (2016): 753–60.

SOBEL, M. "Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models." *Sociological Methodology* 13 (1982): 290–312.

STANFORD HEALTH CARE. "Find a Doctor | Stanford Health Care." 2016. Available at https://stanfordhealthcare.org/search-results.doctors.html

SUNDARARAJAN, V.; T. HENDERSON; C. PERRY; A. MUGGIVAN; H. QUAN; AND W. A. GHALI. "New ICD-10 Version of the Charlson Comorbidity Index Predicted In-Hospital Mortality." *Journal of Clinical Epidemiology* 57 (2004): 1288–94.

THE ECONOMIST. "DocAdvisor." *The Economist*, International Print Edition, July 26, 2014.

Tofighi, D.; and D. MacKinnon. "Monte Carlo Confidence Intervals for Complex Functions of Indirect Effects." *Structural Equation Modeling* 23 (2016): 194–205.

Uber. "How Star Ratings Work." 2020. Available at https://www.uber.com/us/en/drive/basics/how-ratings-work/

Texas Tech University. "Student Evaluations of Courses and Instructors." 2020. Available at https://www.ttu.edu/courseinfo/evals/

Welch, I. "Sequential Sales, Learning, and Cascades." *The Journal of Finance* 47 (1992): 695–732.