Submitted to
manuscript

The Impact of Vertical Integration on Physician Behavior and Healthcare Delivery: Evidence from Gastroenterology Practices

Lina D. Song[1], Soroush Saghafian[2],
Joseph P. Newhouse[2,3,4,5], Mary Beth Landrum[3], John Hsu[6]
[1] School of Management, University College London, London E14 5AA, UK
[2] Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA
[3] Department of Health Care Policy, Harvard Medical School, Boston, MA 02115 USA
[4] Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston MA USA
[5] National Bureau of Economic Research, Cambridge, MA USA
[6] Massachusetts General Hospital, Boston, MA 02114 USA

The U.S. healthcare system is undergoing a period of substantial change, with hospitals purchasing many physician practices ("vertical integration"). In theory, this vertical integration could improve quality by promoting care coordination, but could also worsen it by impacting the care delivery patterns. The evidence quantifying these effects is limited, because of the lack of understanding of how physicians' behaviors alter in response to the changes in financial ownership and incentive structures of the integrated organizations. We study the impact of vertical integration by examining Medicare patients treated by gastroenterologists, a specialty with large outpatient volume, and a recent increase in vertical integration. Using a causal model and large-scale patient-level national panel data that includes 2.6 million patient visits across 5,488 physicians between 2008-2015, we examine changes in various measures of care delivery, including quality, operational efficiency, and spending. We find that physicians reduce provisions of recommended care processes (e.g., anesthesia with deep sedation) after they vertically integrate, and this results in a substantial increase in patients' post-procedure complications. We further provide evidence that the financial incentive structure of the integrated practices is the reason for the changes in physician behavior, since it discourages the integrated practices from allocating expensive resources to relatively unprofitable procedures. We also find that although integration improves operational efficiency (measured by physicians' throughput), it negatively affects quality and the overall spending. Finally, to shed light on potential mechanisms through which policymakers can mitigate the negative consequences of vertical integration, we perform both mediation and cost-effectiveness analyses. Our results indicate that paying about half as much as the current price for a colonoscopy to encourage the provision of deep sedation can be viewed as a cost-effective mechanism to prevent the vertical integration trends from degrading the quality of care.

Key words: Vertical Integration; Healthcare Operations Management; Healthcare Quality; Provider Payment
History: Version: Oct., 2020

1. Introduction

The U.S. healthcare system is undergoing a period of substantial consolidation between physicians and hospitals, with hospitals purchasing many physician practices or directly employing physicians (Kocher and Sahni 2011, Burns et al. 2014, Scott et al. 2017, Baker et al. 2018). The number of physicians who have "vertically integrated" (i.e., consolidated) with hospitals has doubled in the past decade, and the trend is expected to continue (Kocher and Sahni 2011, Neprash et al. 2015, Scott et al. 2017, Nikpay et al. 2018). In theory, there are potential benefits of vertical integration such as greater efficiency through achieving economies of scale and better quality through achieving care coordination and information sharing (Kocher and Sahni 2011, Burns et al. 2014, Carlin et al. 2015, Baker et al. 2018, Baicker and Levy 2013). However, there are also concerns around its anticompetitive effect, which could increase the price and spending and lower quality (Neprash et al. 2015, Capps et al. 2018, Baker et al. 2014). A growing number of studies have examined the impact of integration on care delivery, but the evidence quantifying these effects on quality is still mixed (Post et al. 2018, Scott et al. 2017, Carlin et al. 2015). In particular, a key question that remains unanswered is whether vertical integration promotes a fundamental change in the way physicians operate and how such changes impact various dimensions of care delivery.

In this paper, we generate insights into this question by examining how hospital-physician integration affects quality, efficiency, and spending. We focus on the integration between hospitals and physician outpatient practices, where despite a clear change in the financial ownership of the organization other factors that may affect the care delivery (e.g., physicians, patients, and geographic location) typically remain the same. This enables us to tease out the behavioral responses of the physicians from other contemporaneous changes. We find that when independent physicians integrate with a hospital, they simultaneously reduce recommended care processes (e.g., using deep sedation) and increase their operational efficiency (measured by throughput). The reduction in the process quality, in turn, adversely affects some key dimensions of patient outcomes. We further provide evidence that such behavioral changes are due to the changes in financial incentives of integrated physicians that limit the provision of value-adding care steps after integration. In addition to negative effects on care quality, integration results in an increase in per physician spending. Overall, our results suggest that the changes in financial ownership, without appropriate changes in the incentive structure to motivate the physicians' care processes in a positive direction, can have negative impacts on the healthcare delivery system through an increase in both adverse patient outcomes and spending.

1.1. Policy Context and Setting

Our study setting is the fee-for-service (FFS) Medicare, national health insurance that covers the majority of the elderly U.S. population.[1] Under the FFS payment model, most services are not bundled, and providers are paid for each service at the administratively set prices. Importantly,

[1] In particular, the FFS provider payment policy relevant to the vertical integration we study is Medicare's payment for outpatient services (outpatient prospective payment system).

Medicare reimburses the same procedure in different settings differently: in general, a procedure that occurs in a hospital outpatient department (HOPDs) is paid more than if it took place in a physician's office or an ambulatory surgery center (ASCs). For example, Medicare reimburses $917 on average for colonoscopies that occur in HOPDs in 2019, but only $413 for those in physician offices. Yet, FFS Medicare patients can receive many of the same outpatient procedures in different settings, and there is limited evidence that patients select into different settings to justify such price differentials. As a result, these policies have been criticized for accelerating integration and contributing to the growth in Medicare spending by motivating hospitals to acquire physician-owned practices and convert those practices (usually in the same location) into an HOPD (Office of Inspector General 2014, Medicare Payment Advisory Commission (US) 2017, Forlines 2017, United States Government Accountability Office 2015).

Although the Bipartisan Budget Act of 2015 (Section 603) attempted to eliminate the fee differentials for non-grandfathered practices as of January 2017, the more recent 21st Century Cures Act expanded exemptions further. Clear policy recommendations have been difficult, mainly because evidence has been lacking on how vertical integration affected dimensions of care delivery other than expenditure, especially quality and efficiency. We contribute by establishing evidence on the impact of vertical integration on a variety of outcome measures. From this evidence we generate recommendations for policymakers, showing that if financial incentives can alter physician behavior, paying about half as much as the current price for deep sedation can be cost-effective.

1.2. Challenges and Framework

There are several challenges in providing evidence on the effects of vertical integration. First, many of the existing peer-reviewed studies focus on limited geographic areas, making it difficult to reconcile findings of different effects (Carlin et al. 2015, Wagner 2016). We, however, use a 20% sample of all fee-for-service (FFS) Medicare patients between 2008-2015. These data include 3.6 million observations of patient visits provided by 5,488 physicians. We combine the Medicare data with multiple other data sources that allow us to examine outcomes of quality, efficiency, and expenditures. An overview of our data sources is presented in Table 1.

Second, because the majority of acquisitions occur at a small scale (Capps et al. 2017), vertical integration is not easy to identify from survey data. We exploit Medicare payment rules and providers' billing patterns to infer the financial relationship between physicians and hospitals.

Third, measuring changes in outcomes before and after integration in a meaningful way is challenging. To address this challenge, we focus on a homogenous clinical area: Gastroenterology (GI). GI is one of the specialty areas that has experienced a rapid increase in vertical integration (Nikpay et al. 2018). Figure 1 depicts the trend of integration among gastroenterologists based on our

Table 1: Overview of Our Data Sources

Name	Unit	Years
FFS Medicare Claims (Inpatient, Outpatient, Office)	Patient visit	2008-2015
Medicare Beneficiary Summary Files	Patient	2008-2015
CMS Physician and Other Supplier Data	Physician	2012-2015
Area Health Resource Files	County	2012, 2015
CMS State/County/Plan Enrollment Data	County	2008-2015

data, and shows that integration has consistently increased between 2008-2014. In addition to a rapid increase in integration, focusing on GI has clear advantages because colonoscopy—a primary type of endoscopy for colorectal cancer (CRC) screening and diagnosis—has a set of well-validated process quality measures that are sensitive to the physicians' skills and are linked to important long-term patient outcomes such as interval cancer.

Fourth, establishing causality is difficult because physicians' decisions to integrate are not exogenous; physicians who integrate are more likely to benefit from integration. Specifically, decisions to integrate presumptively depend on physicians' ability, strategy, and technology, which in turn can be correlated with their care delivery patterns. Thus, in order to show that vertical integration causes changes in care delivery, one needs to address the inherent differences between the physicians who decide to integrate with hospitals versus those who do not. To do so we take advantage of our panel dataset and use a Difference-In-Differences (DID) fixed effects model that controls for stable unobserved differences between the comparison groups.

Fifth, integration might be driven by market factors other than reimbursement differences such as technology, market demand, insurance structure, and socioeconomic factors, so we link our panel data to other data sources and adjust for a set of relevant covariates. To gain further confidence, we also conduct multiple robustness checks on our assumptions. Finally, we employ mediation analysis to identify the drivers of the changes we observe, which, in turn, allows us to provide clear policy recommendations.

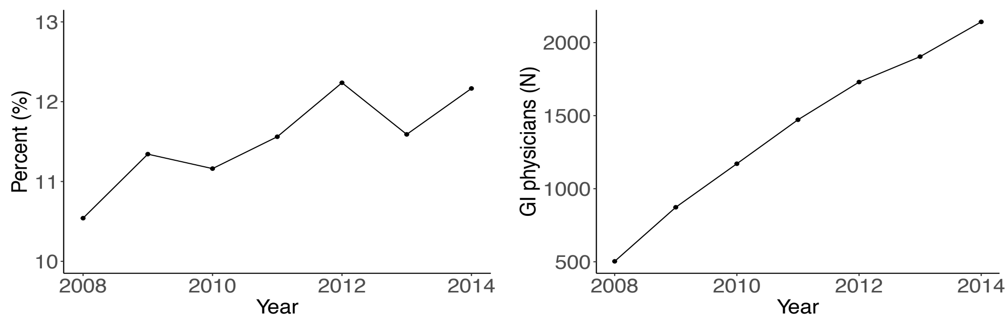
1.3. Main Findings and Contributions

We find that vertical integration negatively affects the quality of care. After physicians integrate, recommended care processes fall, especially the use of deep sedation; about 3.7 fewer patients receive deep sedation per 100 patients.[2] Furthermore, patients of integrated physicians experience a significant increase in both major post-colonoscopy complications such as bleeding, 3.8 per 1,000 colonoscopies, and minor complications such as GI or cardiac symptoms, 3.3 per 1,000 colonoscopies. These effects remain even after adjusting for changes in patient composition and market

[2] Compared to other types of sedation, deep sedation requires more resources and coordination efforts since only anesthesiologists can administer it, whereas other types of sedation can be administrated by nurses. This describes why the impact on deep sedation is higher than other sedation types.

Figure 1: Trend of GI Physician Integration Based on Our Data (2008-2014)

(a) Annual Newly Integrated (b) Cumulative Integrated



Note. From our analysis of Medicare FFS claims. The annual percent of newly integrated GI physicians was calculated as the total number of physicians who newly become integrated in the subsequent year out of the total number of independent physicians in a given year. Cumulative total integrated physicians in a given year was calculated as the sum of all GI physicians who integrated in the previous years and remained integrated in the given year.

characteristics. Through mediation analyses, we find that the reduction in the use of deep sedation, driven mainly by hospitals no longer allocating expensive anesthesiologists to relatively unprofitable colonoscopy procedures, is the main mechanism through which the increase in adverse outcomes occurs. In addition to the decrease in quality, integration increases physicians' throughput and elevates spending per procedure. Notably, we find that integration causes physicians to spend about \$127 more per colonoscopy procedure, which is equivalent to an increase of 48% in mean spending of independent physicians.

Taken together, our results indicate that, despite an increase in spending and operational efficiency, vertical integration does not improve quality. Rather, the shift in the incentive structures of the organization as a result of integration can generate unintended negative consequences in both quality and spending. Our cost-effectiveness analyses reveal that if increasing financial incentives can alter physicians' behavior, paying about half as much as the current price for deep sedation can be cost-effective.

2. Conceptual Framework

2.1. Background in GI Practices

Colonoscopy is a primary screening modality in GI practices for Colorectal Cancer (CRC) screening and diagnosis. During a colonoscopy, a colonoscope—a long, flexible tube—is inserted into the rectum to examine for the presence of polyps. If not removed by the physician during colonoscopy ("polypectomy"), polyps can develop into colorectal cancer. Thus, the key to a good outcome is how well the physician removes any polyps or adenomas, which are benign tumors that may be precursor lesions to CRC. Polyp removal is not always straightforward, however, and because the physician controls the entire process, the outcome is highly dependent on the skill of the physician. There are common post-colonoscopy complications such as bleeding, infection, and in rare cases, perforation of the bowel. Currently, there is a wide unexplained variation in physician quality (e.g., surveillance adherence and adenoma and polyp detection rates) and colonoscopy outcomes (e.g., complications and interval cancer rates) (Corley et al. 2014, Cooper et al. 2012).

2.2. Theories of Vertical Integration

Vertical integration refers to the common ownership of two or more stages of production (or distribution) that initially are separate. Patients can experience a vertical chain of healthcare, including primary care physicians, specialists, hospitals, and rehabilitation facilities. In the colonoscopy case, specialists and hospitals are considered to be the upstream and the downstream entities, respectively. Economic theory suggests two motivations for vertical integration: efficiency-based and strategy-based (Baker et al. 2018, Post et al. 2018). Efficiency-based theories propose that providers integrate primarily to eliminate inefficiencies in production. Strategy-based theories propose that providers integrate to increase market power and/or to employ anticompetitive tactics to create barriers to entry (Gaynor 2014).

The efficiency-based theory claims that if physicians and hospitals are under the same system, cost can be reduced from easier communication, reductions in duplicate services and waste, and goal setting/standardization of practices (Kocher and Sahni 2011, Burns et al. 2014, Baker et al. 2018, Baicker and Levy 2013). This results in clinical integration, which management literature defines as the coordination of patient care services across the various functions, activities, and operating units of a delivery system (Gillies et al. 1993). Although many believe that clinical integration is the gold standard for improving care quality, there is limited direct evidence that vertical integration will actually achieve clinical integration (Singer et al. 2018).

On the other hand, the strategy-based theory suggests that integration for strategic purposes (vertical foreclosure) will have a less direct impact on clinical outcomes, because the principal purpose of integration is to increase an organization's market power and buy referrals. Such actions will not necessarily motivate organizations to achieve any clinical integration, or even worsen care, if increased market power results in lower motivation to compete on quality. While the efficiency-based theory predicts a positive effect, the strategy-based theory predicts a negative one. Because the two theories are not mutually exclusive, the pivotal question centers on the magnitude of changes.

Finally, we note that, if integration changes any fee that an individual physician receives, there can be both income and substitution effects that work in opposite directions with respect to the supply of services, assuming the time per colonoscopy remains constant (McGuire 2000).

3. Literature

Our study is related to the stream of literature that examines service organizations' operational efficiency and quality of service. In particular, it is relevant to studies that examine the role of public policy (e.g., payment policy) and the provider market structures (e.g., mergers, exit, competition) in the operation of service organizations (Chen and Savva 2018, Song and Saghafian 2019). Within the context of vertical integration, most studies focus on their anticompetitive effect, i.e., how integration affects spending and price (Neprash et al. 2015, Baker et al. 2014). More recently, (Vlachy et al. 2017) has shown using a game-theoretic framework that the alignment between the hospital and physicians could have both a positive impact of reducing costs and a negative impact of decreasing quality. Our study focuses on providing empirical evidence of how integration affects the organizations' operational behaviors and quality, as well as spending.

Within the operations management literature that examines worker behavior in service organizations, our study is related to the empirical studies that examine how organizational settings, both financial and non-financial, affect the operational efficiency and quality of services (Tan and Netessine 2019, Wang and Zhou 2018, Meng et al. 2018, Staats et al. 2017). Such studies have examined how specific characteristics of the organization affect worker behavior, such as the structural layout of the facility (Meng et al. 2018, Chan et al. 2019) or the monitoring program (Staats et al. 2017). Other studies have examined the role of innovative payment policies such as the Hospital Readmissions Reduction Program (HRRP) (Arifoğlu et al. 2020, Chen and Savva 2018). Our study examines how the ownership of the organization affects the behavior of workers, and also identifies the specific changes in behavior that can impact performance.

Within the supply chain management literature, our work is related to studies that investigate how vertical integration can improve efficiency by reducing the double marginalization problems. The double marginalization problem has been studied extensively in the operations management literature, mostly through supply chain models (Heese 2007, Li et al. 2013). Fewer studies, however, have empirically evaluated how behavioral changes within the integrating entities may influence the overall effect of vertical integration. Our study contributes by providing an empirical investigation in this regard.

Lastly, our study contributes to medical literature that explores the determinants of medical care quality (Song et al. 2010). Specifically, there are large unknown variations in the quality of GI practices, e.g., in CRC screening and diagnosis (Warren et al. 2009, Rabeneck et al. 2008). Through studying the differences between integrated and not integrated GI practices, our work contributes by shedding light on ways the variations in physician practices can be reduced, thus guiding clinical and public health practitioners.

4. Data and Study Setting

4.1. Data

As noted earlier, Table 1 provides an overview of our data sources. Our main data source is a 20% sample of traditional FFS Medicare claims (Parts A and B) for inpatient, outpatient, and office visits between 2008 and 2015. The FFS Medicare claims provide detailed information on each patient visit, including the procedures received through the Healthcare Common Procedure Coding System (HCPCS) codes, diagnosis through International Classification of Diseases (ICD-9) codes, and spending. We obtained each patient's sociodemographic information such as the age, sex, and 9-digit ZIP code from the Medicare Beneficiary Summary Files (BSF). We obtained each physician's information from the Physician and Other Supplier Public Use File from the Centers for Medicare & Medicaid Services (CMS), which provides information on the characteristics, utilization, and payment information on services and procedures provided to FFS Medicare beneficiaries by physicians (Centers for Medicare & Medicaid Services 2016). We incorporated area level health-care utilization, supply, and sociodemographic information from the Bureau of Health Professions' Area Resource Files (ARF), and the county level penetration rate of Medicare managed care plans from the CMS State/County/Plan Enrollment Data.

4.2. Measuring Vertical Integration

Existing studies have taken at least two different approaches to measure vertical integration, survey-based and claims-based. Many survey-based studies have used data such as the American Hospital Association (AHA) Annual Survey or the SK&A physician survey, both of which include questions on the hospitals' or physicians' relationship with the other (Madison 2004, Cuellar and Gertler 2006, Scott et al. 2017, Baker et al. 2014, Wagner 2016, Capps et al. 2018, Koch et al. 2017). Although survey data can provide a direct source of information on integration, they may miss small integrations, be subject to misclassification, or fail to capture physician-level changes.

The claims-based approaches infer the providers' integration status from their billing patterns (Neprash et al. 2015, Konetzka et al. 2018, Desai and McWilliams 2018, Capps et al. 2018, Clough et al. 2017). Their rationale is that (a) hospital-based providers have a strong financial incentive to report services that occurred at the hospital-owned practices due to the payment differential between the HOPDs and physician offices, and (b) only the practices that are 100% owned by a hospital can bill at the higher HOPD rate.

We use a claim-based approach and take advantage of our detailed data sources to measure integration directly. Specifically, for each physician j in year t, we use our data to first calculate the integration intensity as:

$$INTEG_{jt} = \frac{HOPD_{jt}}{HOPD_{jt} + OFFICE_{jt} + ASC_{jt}}, \tag{1}$$

where $HOPD_{jt}$, $OFFICE_{jt}$, and ASC_{jt} represent the total number of unique HOPD, office, and ASC-based claims, respectively.[3] $INTEG_{jt}$ takes a continuous value between 0 and 1, where $INTEG_{jt} = 1$ indicates physicians who exclusively work at integrated practices ("fully integrated") and $INTEG_{jt} = 0$ indicates physicians who exclusively work at independent practices ("independent"). Based on the distribution of $INTEG_{jt}$ depicted in Figure 2, the majority (79.5%) of physicians in our data set are within the range of $0 < INTEG_{jt} < 1$, which means they are neither independent nor fully integrated but practiced at both independent and hospital-owned practices in a given year ("partially integrated").

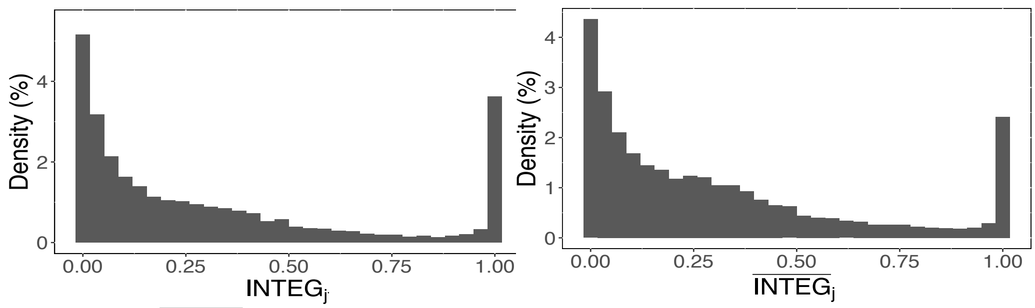
In our main analyses, we define the three types of integration (independent, partial, and full), and make use of 0.1 and 0.9 as upper and lower thresholds of integration intensity in (1). We set the thresholds at 0.1 and 0.9 instead of 0 and 1, because the majority of extremely low integration intensity values (e.g., $0 < INTEG_{jt} < 0.1$) in our data are due to the physicians in the transition stage (e.g., the year that s/he switches from independent to integrated). In our robustness checks, we provide various sensitivity analyses on these thresholds. For example, in addition to re-running our analyses by varying them, we use (1) as a continuous instead of a discrete variable. We also re-run our analysis by using a dichotomized version of integration (see, e.g., Section 9.3 for more details).

Previous studies of integration that have examined other types of physicians, such as primary care physicians (PCPs) or cardiologists, have dichotomized integration status into either independent or integrated (Neprash et al. 2015, Desai and McWilliams 2018, Clough et al. 2017). Dichotomization, however, ignores potentially important differences between partially integrated physicians and both fully integrated or fully independent physicians (Allen and Kaushal 2018). If a hospital simply acquires an independent GI practice, the physicians may just change their integration status from independent to partially integrated. If the physicians are employed by the hospital, however, their integration status changes from independent or partially integrated to fully integrated. Unlike the acquisition scenario, which changes the financial relationship without changes in a physical setting, a change to employment will change the financial and may change the physical environment of the integrating physician. Since we are interested in identifying the impact of financial integration on physician behavior, we mainly focus on the "partial integration" (i.e., when independent physicians become partially integrated), which consists of the majority of the integration cases in our sample.

[3] Unique claims are defined as the claims with the same beneficiary ID, service date, and the provider National Provider Identifier (NPI).

Figure 2: Distribution of $INTEG_{jt}$

(a) In 2015 (b) Average Across 2008-2015



Note. For each physician j, $\overline{INTEG}_j = \Sigma_t INTEG_{jt}/\Sigma_t T_t$ for all years t in the study period for which the physician submitted any claims.

Thus, we eliminate the "full integration" cases in our main analysis, and separately examine their effect later (see Section 8.1).

4.3. Study Population and Comparison Groups

Our patient population is the FFS Medicare beneficiaries who received colonoscopies at any outpatient care settings during our study period, are aged between 65 and 85 at the time of the procedure and are entitled to Medicare due to age.[4] We also focused on the patients who have received colonoscopy from GI physicians.[5] The claims for colonoscopy and related diagnoses were extracted using relevant ICD-9 and HCPCS codes (listed in Table 1 of the Online Appendix).

5. Variable Definitions and Descriptions

5.1. Outcome Variables

We divide our outcome variables into four categories: process-related quality, outcome-related quality, operational efficiency, and spending. Most outcomes were measured from the FFS Medicare inpatient, outpatient, and carrier claims, although some physician efficiency measures (number of services, unique procedures, and patients) were obtained from the CMS Physician and Other Supplier Data. Because of data limitations, these latter three variables were only available for the

[4] The guideline recommends against screening above age 85 (US Preventive Services Task Force 2008). Thus, we removed those above age 85 when they received colonoscopy from our main analysis, as are likely to be clinically different from the rest. We also limited the analysis to those aged 65 or above, since the Medicare beneficiaries under 65 are often sicker than a typical Medicare population. To ensure the observation of post-procedure adverse outcomes, we further restricted our analysis to those who have continuous enrollment in the FFS Medicare Parts A and B of one year before and 30 days after the colonoscopy date.

[5] Although the vast majority of the physicians who perform colonoscopies are GI physicians, a small number of other specialists such as colorectal surgeons and primary care physicians also perform them. Because the other specialists are likely to have different baseline skills, training, as well as patient characteristics, we removed them from our analysis. Gastroenterologists, or GI physicians, were identified by the specialty code on claims (gastroenterology=10). We removed the colonoscopies performed by physicians with specialties such as "Colorectal surgery (=28)", "Internal medicine (=11)", or "family practice (=08)."

years 2012-2015. All other variables were available for the years 2008-2015. Below, we describe each of these separately. A summary of all our outcomes variables is presented in Table 2.

Process-Related Quality. We selected process-related quality measures based on the quality indicators endorsed by professional societies (Rex et al. 2006). From among that group we selected indicators obtainable from the claims data that are (a) widely accepted, and (b) have the potential to be affected by the known variations among GI physicians' practices. For example, studies show that GI physicians with higher polypectomy rates tend to have better patient outcomes such as lower interval cancer (Warren et al. 2009, Kaminski et al. 2010). Thus, we measured the rate of polypectomy of physicians as a proxy for their process-related quality, after adjusting for various patient risk factors. We identified polypectomy rates from claims by the concurrent pathology bills (Warren et al. 2009). Incomplete colonoscopies can also be used as another proxy for measuring process quality, since they can result in missed lesions, a contributor to the interval cancer (Cooper et al. 2012). We obtained incomplete colonoscopies directly from the HCPCS modifier codes 53, 73, or 74 on colonoscopy claims.[6]

A second key process measure of colonoscopy quality that we examined is the method of sedation. Previously, the primary sedation method for screening colonoscopies had been through midazolam and an opioid. More recently, propofol sedation for outpatient colonoscopies has become popular (Khiani et al. 2012). Evidence shows that propofol sedation can provide fast onset of action, short duration of action, amnestic effects, faster recovery and discharge times and increased patient satisfaction (Chen and Rex 2004). Nonetheless, there is a wide variation in the use of propofol for outpatient endoscopy. Thus, we measured the use of deep sedation (i.e., propofol as a method of anesthesia) as another indicator of procedure quality. We used the presence of an anesthesiologist or a nurse anesthetist to identify anesthesiology involvement (Cooper et al. 2012, Khiani et al. 2012).[7]

Outcome-Related Quality. We examined the three most common major complications, perforation, bleeding, and infection, each of which can result in serious disability or death (Rex et al. 2006, Rabeneck et al. 2008). We also measured minor complications, which are defined as the presence of minor GI or cardiac symptoms (Warren et al. 2009). We identified all complications using ICD-9 codes that are present either on or within seven days after colonoscopy. Finally, we measured the

[6] Modifier 53 indicates a discontinued procedure of physician services. Modifier 73 indicates a discontinued HOPD/ASC procedure prior to the administration of anesthesia. Modifier 74 indicates a discontinued HOPD/ASC procedure after the administration of anesthesia. Both modifiers 73 and 74 apply to facility charges.

[7] Because of the FDA regulation, another provider (i.e., an anesthesiologist or nurse anesthetist) must be present during the endoscopic procedure if propofol sedation is used during a colonoscopy. We followed the existing studies that relied on the presence of the CPT-4 code 00810, anesthesia assistance with endoscopic procedure distal to the duodenum, occurring on the same date as the colonoscopy of interest.

Table 2: Outcome Variables and Their Definitions

Category	Variable	Definition
Process-Related Quality	Polypectomy	Removal of at least one polyp during a colonoscopy.
	Incomplete colonoscopy	A colonoscopy that does not evaluate the colon past the distal third of the colon.
	Deep sedation	Use of propofol sedation during colonoscopy.
Outcome-Related Quality	Perforation	Incidence of a hole in the wall of part of the gastrointestinal tract.
	Gastrointestinal bleeding	Major and minor bleeding in the gastrointestinal tract.
	Infection	Incidence of bacterial infections after colonoscopy.
	Minor GI symptoms	Incidence of paralytic ileus, nausea, vomiting, dehydration, abdominal pain, diverticulitis, and enterocolitis.
	Cardiac symptoms	Incidence of arrhythmia, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension, or shock.
	Interval cancer	Incidence of CRC 6 to 36 months after a negative colonoscopy.
Operational Efficiency	Time to complete colonoscopy	Time interval between incomplete colonoscopy to next colonoscopy.
	Time to treatment	Time interval between confirmatory colonoscopy to initiation of cancer treatment.
	Physician efficiency (throughput)	Total number of colonoscopies, services, unique procedures, or patients per GI physician per year.
Spending	Spending per procedure	Total spending occurred during the service event.
	Spending per physician	Total colonoscopy related spending occurred per physician per year.

Note. Major bleeding events include intracranial hemorrhage, hemoperitoneum, and inpatient or emergency department stays for gastrointestinal, hematuria, or not otherwise specified hemorrhage. Minor bleeding events included epistaxis, hemoptysis, vaginal hemorrhage, hemarthrosis and any outpatient claim for hematuria, gastrointestinal, and not otherwise specified hemorrhage.

downstream health outcome, interval cancer, which is CRC that occurs despite receiving a screening colonoscopy. We identified the interval cancer based on existing claims-based approaches (Quantin et al. 2012). We defined a colonoscopy that was received six to 36 months before the diagnosis of cancer as attributable to cancer. Conditional on genetic and clinical variations, interval CRC may occur due to an inadequate polypectomy or missed lesions. Thus, a higher interval cancer rate is a signal of poor physician quality after adjusting for the patient risk factors (Kaminski et al. 2010). The details of measurement and validation can be found in the Online Appendix.

Operational Efficiency. We used physicians' throughput and patients' waiting times as measures of operational efficiency. We calculated throughput in various ways: the total colonoscopies performed per physician per year, the total number of services (i.e., a unique date-physician-provider triplet) provided per physician per year, the total number of unique procedures (i.e., a unique number of HCPCS codes submitted) given per physician per year, and the total number of unique patients treated per physician per year. Next, we measured two waiting times: time from incomplete colonoscopy to the next follow-up colonoscopy, and time from positive colonoscopy to the initiation of surgery.[8] In some cases, a patient may receive initial and follow up colonoscopies from physicians in different organizations, so that the organization responsible for the outcome is unclear. To avoid

[8] For our DID analysis, we limited the sample to the patients who have either received the follow up colonoscopy or surgery within a year of the index colonoscopy. Because of the distributional shape of the time interval variables, we used a logged interval in our analyses.

ambiguity, we limited our analysis to patients who received the two procedures from the same organization.

Spending. Although Medicare unit prices are administratively set, we examined the changes in spending per procedure to determine if integration results in a change in the procedure mix (e.g., treatment intensity) that affects spending. For example, colonoscopy reimbursement rate varies by the type of specific procedure used to remove polyps. Spending per procedure was defined as the total amount paid to the provider per colonoscopy (e.g., a unique date-physician-provider triplet), which we obtained from the claims. To better understand the drivers of overall spending, we also measured annual per physician spending, which is a product of both per procedure spending and the per physician volume (see Table 2).

5.2. Independent Variables

We divide the independent variables used in our models into three categories: patient, physician, and market characteristics. Table 3 has a summary of all independent variables.

Patient Characteristics. We controlled for patient age, gender, race/ethnicity, the reason for Medicare entitlement (i.e., whether or not a beneficiary is entitled to Medicare due to end-stage renal disease, or ESRD), and Medicare-Medicaid dual eligibility status ("Duals"), a proxy for low-income status. We accounted for patients' health risks by calculating each patient's Elixhauser Comorbidity Index (Elixhauser et al. 1998).[9] We also calculated indicators for chronic conditions from the Chronic Conditions Data Warehouse algorithm (Chronic Condition Data Warehouse 2014), which uses diagnosis and procedure codes from the previous year to determine which of 27 chronic conditions the patient may have.

Physician Characteristics. We controlled for the physician's region of practice, the total number of affiliated practices using the same tax identification number (TIN), and an indication of affiliation with a multispecialty clinic, which meant the practice had specialists other than gastroenterology or anesthesiology. Finally, we identified the Ambulatory Surgery Center (ASC) status of each physician's practice based on whether the practice submitted any ASC-based claims.

Market Characteristics. We controlled for the market concentration because horizontal and vertical integration can be correlated. To measure market concentration, we computed Herfindahl-Hirschman Indices (HHIs) for hospitals for each market (HRR). HHIs were calculated by summing the squared market shares of the organizations in the market. We also included the Medicare Advantage (i.e., the managed care type of insurance for Medicare) penetration rates as a proxy for the insurance market structure. We controlled for the provider market supply by including

[9] The Elixhauser Comorbidity Index includes 30 diagnoses that can potentially increase the probability of adverse outcomes. We calculated the index directly from the patient's inpatient and outpatient claims history in the previous year and used the total number of chronic conditions in our main model (Elixhauser et al. 1998).

Table 3: Definition of Independent Variables and Data Sources

Variable	Description	Data source
Patient characteristics		
Age	Numeric, 64-86.	Medicare BSF
Gender	Binary, male or female.	Medicare BSF
Race	Factor, White, Black, Hispanic, Asian, or others.	Medicare BSF
Medicare entitlement	Binary, ESRD or not.	Medicare BSF
Medicaid eligibility	Binary, dual or non-dual.	Medicare BSF
Comorbidity	Numeric, from 0 (least severe) to 21 (most severe).	Medicare inpatient, outpatient claims
Chronic conditions	Numeric, from 0 to 27.	Medicare BSF
Location	Binary, rural or urban.	Medicare Cost Report, POS
Physician characteristics		
Number of affiliations	Numeric, greater than 0	Medicare inpatient, outpatient claims
Multispecialty	Binary, 0 or 1	Medicare inpatient, outpatient claims
ASC affiliation	Binary, 0 or 1	Medicare inpatient, outpatient claims
Market characteristics		
Herfindahl-Hirschman Index	Numeric, greater than 0	Medicare inpatient claims, Medicare Cost Report
Medicare Advantage penetration	Numeric, from 0 (no penetration) to 1 (full penetration).	State/County/Plan Enrollment Data
GI physician density	Numeric, from 0 (none) to 1 (all population) per person	AHRF
Unemployed	Numeric, from 0 (none) to 1 (all population) per county	AHRF
Poverty	Numeric, from 0 (none) to 1 (all population) per county	AHRF
Under age 65	Numeric, from 0 (none) to 1 (all population) per county	AHRF

the total number of GI physicians per person by county from AHRF. Finally, we included county level sociodemographic characteristics: the proportion of the population who are unemployed, in poverty, or are under 65 of age from AHRF.

5.3. Descriptive Statistics

Table 4 summarizes the cross-sectional patient characteristics and the outcome variables by their physicians' integration status in a given year, averaged across 2008-15. The cross-sectional measure of outcome rates is consistent with the estimates from the literature (Rex et al. 2006). Overall, patients who receive a colonoscopy from fully or partially integrated physicians are more likely to be Black, Duals, have a higher comorbidity index, and reside in rural areas (compared to patients receiving treatment from independent physicians). They are also more likely to have higher unadjusted adverse outcomes such as perforation, bleeding, infection, minor complications, and interval CRC. Notably, partially or fully integrated physicians provide deep sedation substantially less, equivalent to about 71% and 48% of the use of the independent physicians, respectively. Such differences between the integrated and independent physicians, as well as their patients, are further examined in our DID analysis, to which we now turn.

6. Main Empirical Strategy

6.1. Overview

Our main empirical strategy is based on a DID analysis with a physician, area, and year fixed effects. Under certain assumptions that we describe in the next section, the coefficient for the

Table 4: Patient Characteristics and Outcome Variables by Integration Status

	Independent	Partial	Integrated
Observations (N)	1,094,303	1,373,297	126,448
Patients (N)	839,145	1,050,834	97,969
Demographic			
Age (mean)	73.18	73.30	73.07
Gender, Male (%)	45.25	44.87	45.65
Race, White (%)	87.29	87.71	84.50
Race, Black (%)	7.04	7.87	10.12
Race, Asian (%)	2.21	1.48	2.05
Race, Hispanic (%)	1.47	1.27	1.15
Duals (%)	8.35	9.07	11.60
Rural (%)	10.12	16.44	20.81
Comorbidity (mean)	1.53	1.59	1.63
Chronic conditions (%)	95.97	95.80	96.69
Process-Related Quality			
Polypectomy (%)	60.15	59.82	61.41
Incomplete colonoscopy (%)	1.66	1.62	2.11
Deep sedation (%)	57.81	41	27.62
Outcome-Related Quality			
Perforation (%)	0.12	0.16	0.20
Bleeding (%)	27.24	29.77	26.71
Infection (%)	0.18	0.25	0.35
Minor complications	11.60	13.48	12.15
Interval CRC (%)	0.24	0.25	0.29
Operational Efficiency			
Total colonoscopies per year (N/physician/year)	180.66	170.32	83.98
Total services per year (N/physician/year)	1,421.67	1,110.09	481.05
Total procedure types per year (N/physician/year)	38.71	42.17	32.87
Total patients seen year (N/physician/year)	454.69	458.96	261.86
Median time to followup (days)	24	28	23
Median time to surgery (days)	30.83	31.83	32.83
Spending			
Provider spending per colonoscopy (USD)	210.95	195.77	192.78
Facility spending per colonoscopy (USD)	388.23	570.52	645.96
Total spending per colonoscopy (USD)	262.81	544.64	742.90
Annual colonoscopy spending per physician (USD)	8,067.09	15,752.99	10,369.53

Note. Patients' integration status is assigned based on the physician they received colonoscopy from. Independent physicians have $0 < INTEG_{jt} < 0.1$, partial physicians have $0.1 < INTEG_{jt} < 0.9$, and integrated physicians have $0.9 < INTEG_{jt} < 1$. All characteristics differed at the significance level 0.001. The average total spending per colonoscopy is smaller than the sum of the average provider and facility spending, because the physician does not receive the facility spending when s/he practices at non-integrated setting. Thus, for most independent and partially integrated physicians, the total spending is determined by the physician and not facility spending.

treatment variable in our model can provide a causal interpretation of how vertical integration affects care delivery. The unit of analysis in our model is a colonoscopy, and as noted above, various characteristics of patients, physicians, and markets are used as controls. The treatment status/variable in our setting is based on the integration measure variable of the physician who performs the procedure. We allow multiple colonoscopies performed on the same patient to have different treatment status if the patient received multiple colonoscopies from different physicians.

The majority (73.0%) of patients, however, received only one colonoscopy during our study period.[10] To perform our DID analysis, we made use of the following model:

$$Y_{ijt} = \alpha POST_{jt} + \beta \mathbf{X_{ijt}} + \gamma \mathbf{Z_{it}} + PHYSICIAN_j + MARKET_i + YEAR_t + \epsilon_{ijt}, \qquad (2)$$

where Y represents outcome variables such as process-related quality, outcome-related quality, operational efficiency, or spending, $POST$ is a binary variable that indicates that the observation is post-integration for the treated group, $PHYSICIAN$ is the physician fixed effect, $MARKET$ is the market fixed effect, and $YEAR$ is the year fixed effect. $\mathbf{X}$ is the vector of patient characteristics, $\mathbf{Z}$ is the vector of market characteristics, and ϵ is an error term. Indices i, j, and t represent a patient, physician, and year, respectively. Bold notation is used to represent vectors. Standard errors are clustered on physician group and year.

6.2. Main Assumptions

The main assumptions of our identification strategy that are needed to support a causal interpretation are: (a) all effects other than integration affect physicians equally, as tested by parallel trends in outcome variables between the treatment and the control group in the pre-integration period, and (b) strict exogeneity. Figure 1 of the Online Appendix confirms that there are similar trends in outcomes before the physicians integrate. For relatively rare outcomes such as incomplete colonoscopy, perforation, infection, and interval cancer, however, the parallel trend is less stable, likely because of the small number of observations. But for outcomes that show significant changes based on our DID analysis such as deep sedation, bleeding, and minor complications, the parallel trend is stable. We also statistically test the differences in trend between the two groups by interacting each of the pre-integration years with our treatment variable (see Table 16 of the Online Appendix). None of the interaction terms are significant, suggesting that there is no significant difference in time trends between the comparison groups prior to integration.

The strict exogeneity condition assumes that the regressors are uncorrelated with the error terms. Such an assumption can be violated if, for example, the errors are correlated with unobserved, time-varying characteristics. Our rich set of covariates for patient and area characteristics, as well as the fixed effects at multiple levels, address the heterogeneity between the comparison groups and year-specific shocks. However, there are still three threats to the strict exogeneity assumption: patient selection of physicians, physician selection into integration status, and changes in physician hidden behavior post-integration.

[10] Among the 385,901 patients who received multiple colonoscopies, 166,918 (43%) received them from the same physician.

Patient Selection to Physicians. Patients may select physicians in a non-random way that is unobservable to us. For example, when physicians join a hospital or a large healthcare organization, patients' perception of the quality of the service may change such that poorer or sicker patients select into integrated physicians. This can make it appear that integration worsens quality. Indeed, as noted before, we observe some baseline differences in patient characteristics in our data (Table 4). Because the selection has to be both time-varying and unobservable, this does not seem like a major threat to us. This concern is further mitigated, since we adjust for overall comorbidity, chronic conditions, and sociodemographic status. Also, many patients do not choose their specialists directly but are often referred by their PCPs (Barnett et al. 2012). Nevertheless, we later examine the changes in observed patient risk composition before and after the physicians integrate to further address this concern.

Physician Selection to Integration Status. As noted above, physicians' decisions to integrate are unlikely to be exogenous, and might result from strategic behavior. For example, those who decide to integrate may prefer to collaborate with others more, which may also be correlated with their quality. It is also possible that the physicians' decisions to integrate are based on the pre-integration characteristics of their patient group. For example, physicians with a greater proportion of low-income or high-risk patients may decide to integrate to alleviate financial struggles. Finally, the market characteristics such as the degree of horizontal integration, the degree of managed care penetration, or the input costs may be correlated with both physicians' propensity to integrate and underlying patient health (Gaynor et al. 2013). Because these threats only apply to an extent the differences are unobservable and time-varying, they should be mitigated by the various controls we include. We further examine the physician selection effect in multiple ways, including examining the effect of market conditions.

Physician Hidden Behavior Post-Integration. Integrated physicians may change their behavior in a way that confounds the integration effect. For example, physicians affiliated with an integrated organization are more likely to increase the coding intensity for reporting complications, which may affect the outcome-related quality measures we examine without impacting the true underlying quality. In particular, if physicians code for complications after colonoscopy more actively after they integrate due to increased monitoring efforts, it may appear that the quality has worsened post-integration. However, we believe that this concern is mitigated for several reasons. First, major complications are less likely to be subject to variations in this coding behavior than minor complications, but our results show a stronger effect for major complications (e.g., bleeding). Second, we account for patients' ability to visit any of the inpatient, outpatient, or office settings for subsequent adverse outcomes, and are not limited to the same practice that they originally visited. This adverse outcome is, however, attributed to the original practice where they received a colonoscopy, not the practice they visit with the complications. Third, our examination of patient composition shows that the proportion of high-risk patients is fairly consistent pre- and post-integration, which weakens the argument that providers increase their coding intensity. Nevertheless, to gain further confidence, in our robustness checks, we test these assumptions in various ways (see Section 9).

7. Results and Discussion

7.1. Average Effects of Integration

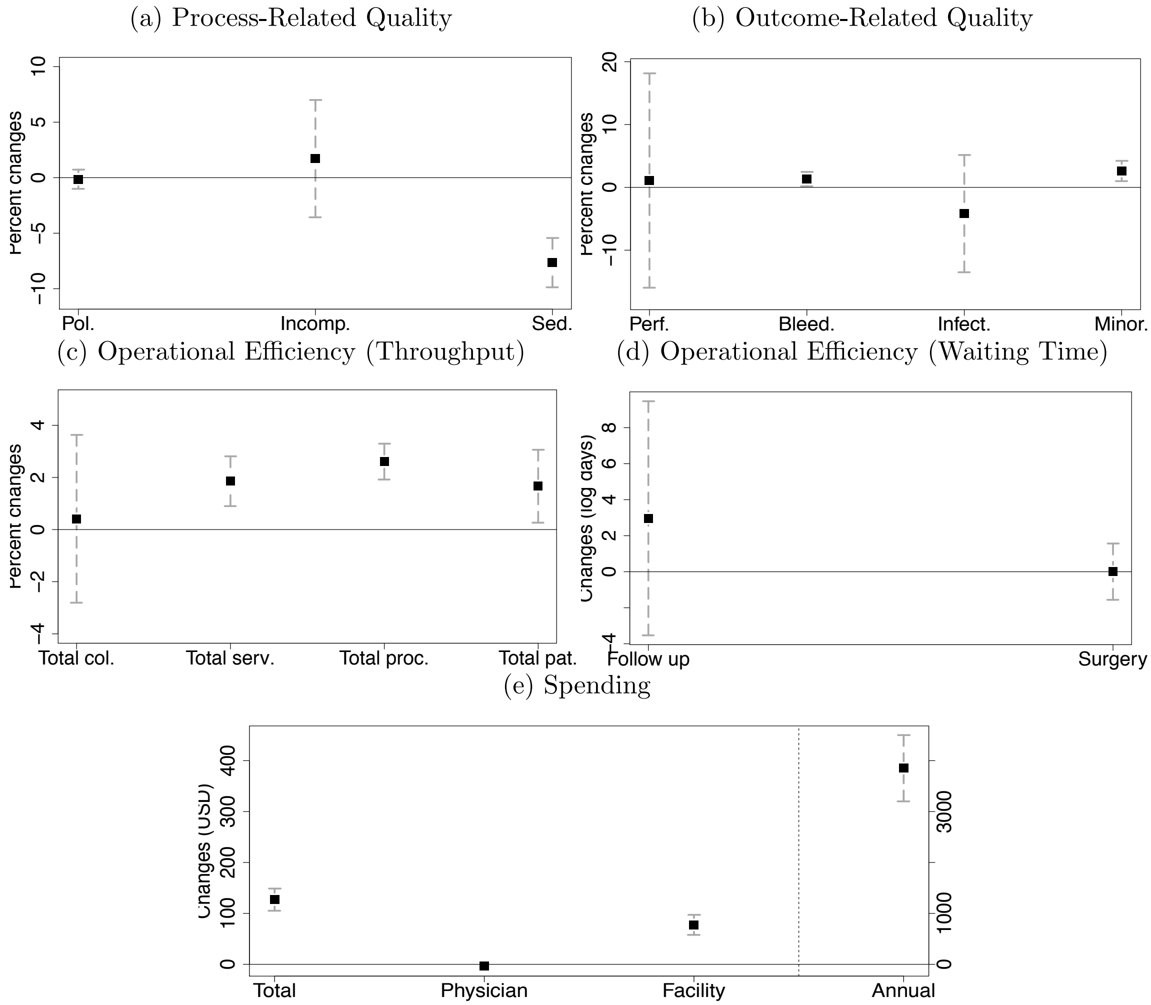
As noted earlier, in our main analysis we focus on the impact of independent physicians becoming partially integrated, which constitutes the majority of integration cases in our data. Thus, in what follows, we simply label this type of integration as "integration." In later sections we expand our analyses to other types of integration (see Section 8.1), and provide various robustness checks on our definitions of integration types (see Section 9.3).

Process-Related Quality. Figure 3 and Table 5 show the estimates of our DID coefficients. Full regression results are provided in the Online Appendix (Tables 2-5). Among the three process-related quality (polypectomy, incomplete colonoscopy, and deep sedation), polypectomy and incomplete colonoscopy do not change after integration. However, the physicians who integrate reduce the use of deep sedation by 7.7%, equivalent to about 3.7 fewer patients receiving deep sedation per 100 patients receiving colonoscopies.

Outcome-Related Quality. Among the outcome-related quality (perforation, bleeding, infection, minor complications, and interval cancer), we observe that patients experience a significant increase in bleeding and minor complications after colonoscopy when their physicians integrate. Because both bleeding and minor complications are relatively common complications (e.g., the average 30-day incidence of bleeding and minor complications are 28.6% and 12.6%, respectively), such increases translate into about 3.8 and 3.3 additional bleeding and minor complications out of 1,000 colonoscopies, respectively. We have also conducted multiple testing adjustments for the three process- or outcome-related quality measures that are significant under the assumption of independent hypothesis: deep sedation use, bleeding, and minor complication. We used Bonferroni, the most conservative among multiple comparison tests (Benjamini and Hochberg 1995), and the family wide error rates remain significant at < 0.001, < 0.01, and < 0.05, respectively (see Section 9.5).

Operational Efficiency. Overall, when GI physicians integrate, there are no significant changes in the waiting time, either to a follow-up colonoscopy or surgery after a positive colonoscopy (Figure 3). However, there are noticeable changes in the throughput measures: the GI physicians

Figure 3: DID Estimates: Average Efect of Integration



Note. Pol. indicates polypectomy. Incomp. indicates incomplete colonoscopy. Sed. indicates deep sedation. Perf. indicates perforation. Bleed. indicates bleeding. Infect. indicates infection. Minor. indicates minor complications. Col. indicates colonoscopy. Serv. indicates services. Proc. indicates procedures. Each dot indicates the size of the DID coefficient. Each dot indicates the size of the DID coefficient. Grey lines depict the 95% confidence intervals around the coefficient of the DID variable. Standard errors are robust and clustered at the physician and the year levels.

significantly increase three out of four measures of throughput (services, procedure types, and patients) after integration.

Spending. Our results show that for each colonoscopy visit, a physician's integration is associated with a $127 increase in total Medicare spending for a colonoscopy (Figure 3). The effect is driven by an increase in the facility fees after a physician has integrated. The estimated change in spending is consistent with the price differential between physician offices and HOPDs, suggesting that integrated physicians do not alter their procedure mix (e.g., by increasing the provision of cheaper polypectomy methods and reducing more expensive ones). We further verified that there are no significant changes in the proportion of specific procedure types (e.g., single or multiple biopsies, or

Table 5: Regression Results

	Dependent variable: process and outcome-related quality							
	Pol.	Incomp.	Deep Sed.	Perf.	Bleed.	Infect.	Minor Comp.	Int. CRC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
POST	−0.00084	0.00028	-0.037^{***}	0.00002	0.0038^{**}	−0.00009	0.0033^{*}	−0.00014
	(0.0017)	(0.00044)	(0.0059)	(0.00012)	(0.0012)	(0.00010)	(0.0010)	(0.00026)
Obs.	2,442,582	2,442,582	2,442,582	2,442,582	2,442,582	2,442,582	2,442,582	1,551,886
R^2	0.097	0.032	0.529	0.013	0.127	0.022	0.102	0.019
Adj. R^2	0.092	0.026	0.527	0.007	0.122	0.016	0.098	0.012
Res. Std. Err.	0.467	0.125	0.344	0.037	0.424	0.047	0.316	0.056

	Dependent variable: operational efficiency and spending									
	Total Serv.	Total Col.	Total Proc.	Total Pat.	Follow. Time	Surg. Time	Total Spend.	Phy. Spend.	Fac. Spend.	Ann. Spend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST	2.287^{*}	5.146	3.241^{**}	1.059^{**}	0.078	0.0002	126.956^{***}	-3.542^{***}	77.477^{***}	$3,851.020^{***}$
	(20.460)	(0.852)	(0.142)	(3.260)	(0.087)	(0.027)	(11.105)	(0.536)	(10.043)	(331.907)
Obs.	24,025	66,908	34,398	33,337	16,124	40,639	2,442,582	2,442,582	976,456	67,362
R^2	0.924	0.871	0.922	0.947	0.641	0.291	0.359	0.263	0.365	0.816
Adj. R^2	0.871	0.845	0.889	0.924	0.504	0.069	0.355	0.259	0.357	0.778
Res. Std. Err.	721.685	52.363	5.739	78.940	1.186	1.148	287.040	60.216	292.382	5,690.967

Note. $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Pol. indicates polypectomy; Incomp. indicates incomplete colonoscopy. Deep Sed. indicates deep sedation; Perf. indicates perforation; Bleed. indicates bleeding. Infect. indicates infection; Minor Comp. indicates minor complication; Int. CRC indicates interval CRC; Serv. indicates services; Col. indicates colonoscopy; Proc. indicates procedures; Pat. indicates patients; Follow. indicates follow-up; Surg. indicates surgery; Spend. indicates spending; Phy. indicates physician; Fac. indicates facility; Ann. indicates annual; Obs. indicates observations; Adj. R^2 indicates adjusted R^2; Res. Std. Err. indicates residual standard error. Standard errors are robust and clustered at the physician and the year levels.

specific types of polypectomy methods[11]) before and after a physician integrated, which suggests that any changes in the procedure mix did not drive up the spending (see Online Appendix Table 6). Annual colonoscopy Medicare spending increases by $3,851 per year after physicians integrate, driven by both an increase in physician throughput and an increase in per procedure spending. Thus, vertical integration increases spending not only through the increase in the administratively set price, as existing evidence has shown, but also by changing the physician behavior to increase their throughput.

In summary, our analysis shows that integration negatively affects some important dimensions of care delivery, including quality and overall spending. Most notably, despite a significant increase in spending, we find that the patients of integrated physicians experience worse outcomes in some quality measures such as rates of bleeding and minor complications. How can policymakers avoid these unintended negative consequences of integration? To answer this question, we next examine the mechanisms behind our results.

[11] Biopsy is defined as a colonoscopy with single or multiple biopsies (CPT 45380). Polypectomies were identified as colonoscopies with ablation (CPT 45383), hot biopsy forceps or bipolar cautery forceps (CPT 45384), or snaring (CPT 45385).

7.2. Mechanisms

Insights from operations management literature suggest that changes in operational processes in the service system, such as increases in service speed and/or customer waiting time, can negatively affect the quality of service (Chan et al. 2016, KC and Terwiesch 2009, Song and Saghafian 2019). Since integrated physicians increase throughput while their patients experience worse outcomes, one potential explanation of the negative effects on patient outcomes might be the increase in the throughput, i.e., speeding up the procedure. A second, not mutually exclusive, potential explanation is the reduction in deep sedation. The medical literature indicates several benefits of deep sedation during colonoscopies, including its fast onset of action, short duration of action, amnestic effects, and faster recovery and discharge times, which in turn can improve quality (Chen and Rex 2004). A direct link between deep sedation rates and post-colonoscopy complications is not established, but the reduction in deep sedation among integrated providers may have resulted in increased patients' complications. To examine this, we first included deep sedation as an independent variable and re-examined the effect of integration on patients' outcomes using the following model:

$$Y_{ijt} = \alpha POST_{jt} + \delta SEDATION_{ijt} + \beta \mathbf{X_{ijt}} + \gamma \mathbf{Z_{it}} + PHYSICIAN_j + MARKET_i + YEAR_t + \epsilon_{ijt}, \tag{3}$$

where $SEDATION_{ijt}$ is the binary variable for whether deep sedation was accompanied, and all variables are as previously defined. For examining the role of throughput, we use the following model:

$$Y_{ijt} = \alpha POST_{jt} + \delta THRU_{jt} + \beta \mathbf{X_{ijt}} + \gamma \mathbf{Z_{it}} + PHYSICIAN_j + MARKET_i + YEAR_t + \epsilon_{ijt}, \tag{4}$$

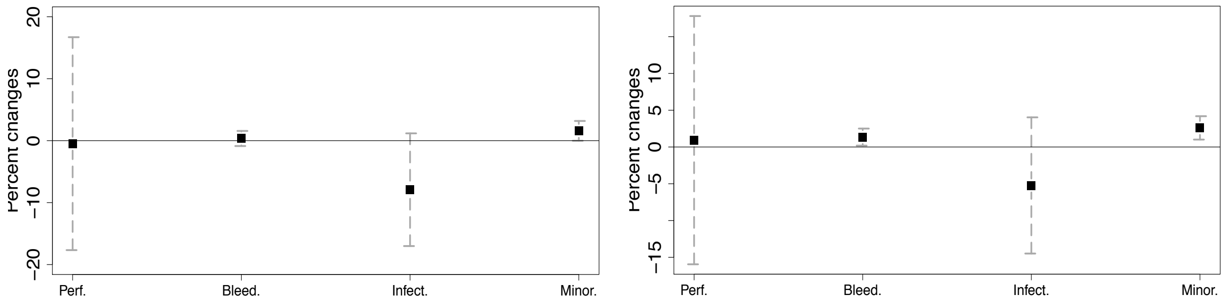
where $THRU_{jt}$ is the colonoscopy throughput per year for physician j in year t.

Our results presented in Figure 4 (and Table 7-8 of the Online Appendix) show that the effect of integration on patients' outcomes is either significantly reduced in magnitude or is no longer statistically significant after adjusting for deep sedation during the procedure, whereas the effect is consistent after adjusting for physicians' throughput. At the same time, the coefficient for deep sedation is significant and negative, indicating that providing deep sedation is associated with a reduction in adverse outcomes. These results suggest that the reduction in the use of deep sedation after integration explains an increase in some of the adverse patient health outcomes.

To further provide support for our claims, we implemented simple mediation models using the R package for causal mediation analysis developed by (Tingley et al. 2014). Using a bootstrapping analysis of 10,000 iterations, we estimated the direct and indirect effects of integration via each of the potential mediators (anesthesia use and throughput) on our outcomes of interest (bleeding and

Figure 4: DID Estimates: Mediation Analyses

(a) Integration Adjusting for Deep Sedation (b) Integration Adjusting for Physician Efficiency



Note. Minor comp. indicates minor complications. All effects are scaled as changes in percentages. Each dot indicates the size of the DID coefficient. Grey lines depict the 95% confidence intervals around the coefficient of the DID variable. Standard errors are robust and clustered at the physician and the year levels.

minor complications). The results in Table 6 confirm that integration had a significant effect (i.e., the 95% confidence interval for the size of indirect effect excludes zero) on anesthesia use, which in turn had a significant impact on both bleeding and minor complications. When we examined the mediation effect of throughput, however, we observed reductions for both bleeding and minor complications, suggesting that efficiency does not mediate the effect of integration on these adverse outcomes.

Given our finding that the reduction in deep sedation use post-integration results in adverse patient outcomes, we next take a closer look at the potential drivers of the changes in deep sedation itself. Compared to other types of sedation, deep sedation requires more resources and coordination effort because only anesthesiologists can administer it, whereas other types of sedation can be administrated by nurses. Thus, the provision of deep sedation is sensitive to anesthesiologists' availability within the organization. There are two potential channels through which anesthesiologists' availability post-integration might affect the care provided to the patients: (1) changes in the external margin (e.g., fewer total anesthesiologists), and (2) changes in the internal margin (e.g., less provision of deep sedation for colonoscopy per anesthesiologist). An example of the first channel is an integrated practice only using its own smaller number of anesthesiologists. An example of the second is an integrated practice shifting anesthesiologist volume to procedures other than colonoscopies.

To examine the first potential channel, we tested whether an integrated GI physician experiences a reduction in the total number of anesthesiologists or nurse anesthetists that s/he works with. To this end, for each GI physician, we first measured the total number of anesthesiologists who have worked with a GI physician in a given year.[12] We found that, on average, a typical GI physician

[12] We include the anesthesia administered by either anesthesiologist (CMS provider specialty code = 05) or anesthesiologist assistants/Certified Registered Nurse Anesthetists (CRNA) (CMS provider specialty code = 32). A GI

Table 6: Mediation Analysis Results

Mediator Outcomes	Anesthesia Bleeding	Minor comp.	Throughput Bleeding	Minor comp.
Average causal mediation effect	0.0037*** (0.0035-0.00)	0.00059*** (0.00046-0.00)	-00029* (-0.00046-0.00)	-0.00015* (-0.00027-0.00)
Average direct effect	0.015*** (0.014-0.02)	0.016*** (0.015-0.02)	0.020*** (0.019-0.02)	0.017*** (0.016-0.02)
Total Effect	0.019*** (0.0180-0.02)	0.017*** (0.016-0.02)	0.020*** (0.018-0.02)	0.017*** (0.016-0.02)
Prop. Mediated	0.19*** (0.18-0.21)	0.034*** (0.028-0.04)	-0.015* (-0.023–0.01)	-0.0091* (-0.016-0.00)

Note. *p<0.05; **p<0.01; ***p<0.001. Comp. indicates complications. Prop. indicates proportion.

works with 6.6 anesthesiologists or nurse anesthetists per year. We next used a DID model similar to our main model but with the total number of anesthesiologists an integrated GI physician works with per year as the outcome variable to identify whether the change in integration affects this outcome variable. Instead of a reduction in the number of affiliated anesthesiologists, we observe a slight increase among integrated physicians (DID coefficient: 0.81, SD: 0.12, p-value: < 0.001), which suggests that the first channel is unlikely to drive the reduction in deep sedation use.

To examine the second channel, we measured the change in the rate of deep sedation exclusively for colonoscopy performed by an integrated versus an independent anesthesiologist. We used a physician level DID analysis where the outcome is the number of deep sedation for colonoscopy per anesthesiologist per year, adjusting for the anesthesiologist and year fixed effects. We find that after an anesthesiologist integrates, s/he provides deep sedation for 2.79 (SD: 0.22, p-value: $<$ 0.001) fewer colonoscopies per year (see Online Appendix Table 9-10 for more details). Thus, these results suggest that while integrated practices do not necessarily reduce the number of the affiliated anesthesiologists, they shift their use of anesthesiologists to services other than colonoscopies. But why do they so?

To answer this question, we hypothesize two explanations. First, providers may shift the allocation of anesthesiologists because of immediate financial gains from the FFS payment differential. This can happen if the immediate gains in providing anesthesia (e.g., anesthesiologists' reimbursement rate) for other procedures is higher than that of colonoscopy. However, this is unlikely because Medicare pays anesthesiologists as a function of their time spent in the operating room. Thus, there is little incentive for allocating anesthesia to different types of procedures to receive a higher payment. Furthermore, the average FFS Medicare payment for deep sedation per GI procedure via anesthesiologist involvement in HOPDs is generally low (only $157.3 in 2012).

Second, integrated providers may shift the allocation of anesthesiologists for longer-term financial gains. When the supply of anesthesiologists (a relatively expensive type of provider) is rigid,

physician and an anesthesiologist or a nurse anesthetists were considered to have worked together if they treat the same patient on the same date.

and there are other specialty procedures that are more profitable (e.g., orthopedics or pain management), the marginal benefit of providing deep sedation for colonoscopy procedures through anesthesiologist may be low. Thus, organizations may allocate anesthesiologists to the services that generate greater overall revenue. We examined the heterogeneity in the reduction of anesthesia among the practices with and without other specialties and find that physicians who integrate with a multispecialty clinic reduce deep sedation about twice as much (-0.037, SD: 0.0069) as those who integrate with a single specialty clinic (-0.019, SD: 0.0097) (see Online Appendix Table 11 for more details). Thus, our results imply that this second mechanism is more likely to be the reason behind the shift in the use of anesthesiologists to services other than colonoscopies. We do not have access to data to directly establish a casual relationship in this regard and leave it to future research do examine this more rigorously. However, our conversation with gastroenterologists also confirms that the related scheduling and administrative processes for accessing anesthesiologists for colonoscopies are challenging because patients undergoing other procedures compete for their availability. Put together, all of our results suggest that modifying the underlying incentive structure can mitigate or prevent the adverse impacts of vertical integration. We discuss this in further detail in Section 10.

8. Heterogeneous Effects

8.1. Examination of Full Integration

In our main analysis, we focused on the integration among the physicians who change status from independent to partially integrated (the majority of integration cases among GI physicians in our data). Here, we separately examine the cases when partially integrated physicians become fully integrated. We do so by applying the same approach and model specification we used for our main analysis (our results for this case are presented in the Online Appendix; see Table 12 there). Unlike the partially integrated physicians in our main analysis who reduce the deep sedation use and increase some of their patients' post-colonoscopy complications, fully integrated physicians do not reduce the use of deep sedation, nor do their patients experience any increase in complications. Moreover, despite experiencing an increase of $75.4 in per procedure spending driven by the administratively set price differentials, fully integrated physicians decrease their throughput after integration.

In summary, our results confirm that the behaviors of the fully integrated GI physicians are likely driven by different motivations than the ones affecting the majority of integrated GI physicians, the focus of our main analysis. One potential reason is that a large proportion of full integration cases involves the physicians becoming hired into hospital-based outpatient practices. This will likely involve a different payment scheme as well as changes in the work environment than the integration cases through acquisitions we study in our main analysis. However, our data is insufficient to rigorously test these and other possibilities, and we leave it to future research to shed further light on these issues.

9. Robustness Checks

To test the spread and validity of our results, we performed robustness checks on various factors that can affect our results. As we describe next, our robustness checks include testing for changes in patient risk composition, investigating physicians' behavior (e.g., coding and gaming or retirement propensity), changing the measure of integration, examining the role of confounders such as the market competition, conducting statistical tests such as Bonferroni correction (to address the risk of having positive results by chance when we conduct multiple comparisons on different outcomes), and inverse probability weighting to adjust for the baseline differences in physicians. We also examined the impact of integration on patient experience to test whether integration affects other dimensions of care using Outpatient Consumer Assessment of Healthcare Providers and Systems (OAS CAHPS) survey data for HOPDs (see Table 15 in the Online Appendix). The specific methods and results of these tests are provided in the Online Appendix. Overall, our results give us confidence that our results are fairly robust and are not sensitive to our assumptions and model specifications.

9.1. Changes in Patient Risk Composition

One important assumption of our identification strategy is that the changes in the quality outcomes of physicians who alter their integration are not due to the changes in patient characteristics as a result of post-integration selection. Given that distance is one of the primary factors for patients' choice of physicians and the majority of physicians' physical locations do not change after integration, such concerns on patient selection before and after integration are likely mitigated. However, it is possible that the changes in the ownership status of a physician can result in attracting a different set of patients. If this happens, then most likely there will be changes in some of the observable composition of patients. Thus, by examining whether there are any changes in observable patient characteristics, we can also get a better understanding of whether the unobservable changes may be a significant threat to our identification strategy. To this end, we first examined the changes in observed patient composition when GI physicians integrate. For each physician, we tested whether the composition of his/her patients with respect to certain characteristics (e.g., demographic, clinical) changed following integration using a physician-level DID model:

$$Y_{jt} = \alpha INTEG_{jt} + \beta \mathbf{X_{jt}} + \gamma \mathbf{Z_{jt}} + PHYSICIAN_j + YEAR_t + \epsilon_{jt}, \tag{5}$$

where Y_{jt} is a measure of patient composition for physician j in year t (e.g., percentage of physician j's patients in that year that have certain characteristics). Other variables are defined the same as our main model.

Figure 5 shows that integration was associated with some changes in demographic or clinical composition. In particular, integrated physicians face a significant increase in the proportion of "Dual" patients and a reduction in the proportion of patients with high risk for CRC.[13] Of note, the DID model used in our main analysis controls for such observable changes. Yet, such changes suggest that potential unadjusted confounders might be present. Empirical evidence shows that patients' risk for CRC is the strong predictor of colonoscopy outcomes (Johnson et al. 2013). Thus, if the unobserved changes in patient characteristics also alter in the same direction as the CRC risk after integration, the result suggests that the changes in patient composition would have biased our results in a direction that underestimates the negative impact of integration on patients' colonoscopy outcomes. That is, our estimates on the magnitude of the negative effect of integration on patients' outcomes might be conservative (i.e., the actual negative impact might be worse than what our estimates suggest). Thus, our robustness checks give us confidence about the direction of our results: integration negatively affects patients' outcomes.

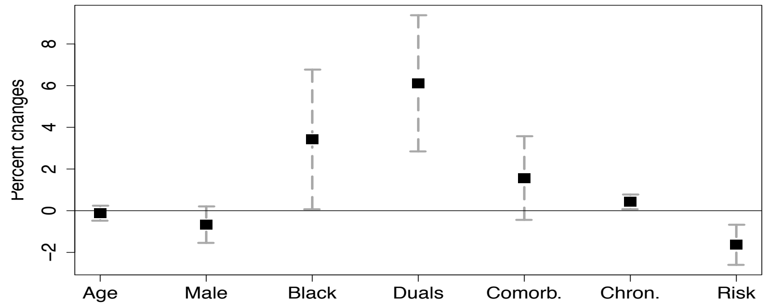
9.2. Physician Behavior

Coding and Gaming. Another important threat to our identification strategy is that physicians may change their coding behavior in a way that does not reflect the true changes in quality, depending on the administrative infrastructure of the newly integrated system. To test this assumption, we made use of the primary condition only to re-examine the integration effect on the two outcomes, bleeding and minor complications. Next, we examined limiting the definition of bleeding to major bleeding only, which are less likely to be subject to change as a result of changes in coding intensity. Our results in Table 7 suggest that there is no evidence that the potential changes in the coding intensity after integration would affect our main findings. Further, there is no good reason for physicians to change their coding behavior for the purpose of gaming, because the measures we use to detect changes (e.g., bleeding and minor complications) are not used for payment or other factors that can create specific incentives.

Retiring Physicians. We also tested if the shift to employment caused GI physicians to become part-time. For example, if the physicians who choose to integrate intend to do so for different reasons (e.g., on a path to retirement), this may affect both the efficiency and quality of their

[13] Medicare considers an individual at high risk if s/he has one or more of the following: a close relative who has had colorectal cancer or an adenomatous polyp; a family history of familial adenomatous polyposis; a family history of hereditary nonpolyposis colorectal cancer; and a personal history of adenomatous polyps, colorectal cancer, or inflammatory bowel disease, including Crohn's disease, and ulcerative colitis.

Figure 5: Average Effect of Integration on Patient Composition, Scaled into Percent Changes



Note. All effects are scaled as changes in percentages. Each dot indicates the size of the DID coefficient. Grey lines depict the 95% confidence intervals around the coefficient of the DID variable. Standard errors are robust and clustered at the physician and the year levels.

Table 7: Robustness Checks Results

	Outcomes	Deep sedation	Bleeding	Minor complications
Behavior	Major bleeding	NA‡	0.0038† (0.0017)	NA‡
	Exclude retiring physicians	-0.037^{***} (0.006)	0.004^{*} (0.002)	0.003^{*} (0.001)
Threshold	Cutoffs at 1%, 99%	-0.015^{*} (0.005)	0.004^{*} (0.002)	0.001 (0.002)
	Cutoffs at 5%, 95%	-0.041^{**} (0.008)	0.001† (0.002)	0.004^{*} (0.001)
	Cutoffs at 15%, 85%	-0.039^{***} (0.007)	0.003^{*} (0.001)	0.006^{*} (0.002)
	Binary integration variable	-0.039^{***} (0.0057)	0.0050^{*} (0.0016)	0.0041^{**} (0.0011)
	Continuous integration variable	-0.139^{***} (0.020)	0.011 (0.007)	0.001† (0.001)
Competition	Low	-0.031^{*} (0.009)	0.005^{*} (0.002)	0.004^{*} (0.001)
	High	-0.041^{***} (0.005)	0.004† (0.002)	0.003† (0.002)
IPW		-0.030^{***} (0.006)	0.011^{*} (0.003)	0.006† (0.003)

Note. † indicates marginally significant at p-value < 0.10. ‡ indicates the results are not subject to change. We only present the results for the outcomes that had any significant changes in the main analyses. Standard errors in parentheses are robust and clustered at the physician and the year levels.

care differently from others. We first identified the GI physicians who are highly likely to retire as those who have submitted any claims for at least subsequent years but submitted no claims (including inpatient claims) for all subsequent years. We identified a total of 1,131 GI physicians (6.9% of total GI physicians in our data) who are likely to be on a path to retirement during our observation period and re-ran the analyses by focusing on them. We observe that these physicians' changes in throughput and the use of deep sedation after integration are similar to the non-retiring physicians, and that our main results are consistent when the retiring physicians and their patients are removed from the sample (see Table 7).

9.3. Measuring Integration

In our main analysis, we made use of specific threshold values on the integration intensity measure introduced in (1) (10% and 90% for partial and full integration, respectively) to define integration. Although these thresholds are supported both by our data and the literature on the practice patterns of GI physicians, we tested the sensitivity of our findings to these values. Specifically, we made use of the following alternative thresholds: 1% and 99%, 5% and 95%, and 15% and 85% for partial and full integration, respectively. Table 7 shows that varying the threshold does not affect our main findings. We also used the binary integration variable (independent vs. integrated) instead to examine the integration effect. This is equivalent to considering all of the physicians who are either partially or fully integrated simply as "integrated". Defining integration in such a way yielded integration effects similar to that of partial integration in our main analysis, likely because the majority of integration cases in our data are related to partial integration, as we noted earlier. Finally, as another robustness check, we directly used the continuous integration measure defined in Equation (1) without imposing any threshold. This also showed consistent results with our main findings.

9.4. Role of Competition

Integration can contribute to reduced competition in the market. Given the fixed Medicare price, reduced competition might incentivize the practices to reduce their efforts on improving quality and/or efficiency. Thus, we examined whether the level of competition in the market plays a role in the integration effect we observe by stratifying our sample into equal sizes of high versus low competition areas. We defined high and low competition areas as those that have higher than median and lower than median Hirschman-Herfindahl Indices (HHIs), respectively. We do not observe noticeable heterogeneous effects between these two groups (Table 7). This suggests that if unobserved changes in the market structure are consistent with the changes in market competition, then market competition is most likely not a major driver of our main findings.

9.5. Other Statistical Tests

Some of our outcome measures are compound outcomes that are potentially correlated with each other, and there is a risk of having positive results by chance when we conduct multiple comparisons on different outcomes. For example, perforation can result in increased bleeding. Hence, we corrected for the potential correlation between different outcomes using Bonferroni and other similar tests.[14] The results of the adjusted outcomes are still significant or marginally significant for all our quality outcomes (see Table 13 in the Online Appendix). Next, we further adjusted for the baseline differences in physicians and patients using inverse probability weighting (IPW) (see

[14] We also applied several approaches to adjust for p-values, including Hochberg, Hommel, Holm, and Benjamini Hochberg and Yekutieli procedures. These approaches vary in the conservativeness of how many p-values they adjust.

Table 14 in the Online Appendix). [15] The results in Table 7 show that our main findings still hold with slightly stronger effects on variables related to adverse outcomes.

10. Policy Implications

Our results overall paint a negative picture of vertical integration, because it decreases some aspects of quality and increases spending. Such evidence, however, does not necessarily indicate that the trend of vertical integration should be reversed. Indeed, we find that vertical integration also brings in potentially positive effects, such as an increase in physicians' throughput. Our evidence also suggests that the negative impact of vertical integration is driven by the physicians' responses to the misaligned financial incentives, rather than other aspects of integration itself (e.g., increased coordination or volume). Thus, one immediate solution is to fix the current payment structure of integrated practices in a direction that promotes better quality. For example, integrated physicians could improve their care delivery process and patient outcomes if the payment for providing deep sedation among integrated practices is adjusted such that it is more consistent with the opportunity costs.

To assist policymakers, we performed counterfactual analyses to estimate the adverse outcomes averted under the scenario of the provision of deep sedation. If all GI physicians who did not provide deep sedation had done so, about 26 bleeding and 12 minor complications per 1,000 colonoscopies would have been averted.[16] Using our estimate, we calculated the reasonable amount of incentive that can be provided to promote the provision of deep sedation based on an existing approach of cost-effectiveness anaalysis (Pandya et al. 2020). We first estimated the health gains, or quality-adjusted life years (QALYs), resulting from increased use of deep sedation as follows. Suppose we can prevent the reduction of deep sedation use after integration through increasing the incentives. Assuming 0.1 QALY the maximum amount of utility loss associated with the post-colonoscopy adverse outcomes based on the literature (Graves et al. 2007) and using a figure that spending up to $22,289 for a unit of QALY is considered by some to be cost-effective in developed countries (Bertram et al. 2016), we back-calculated the monetary level of acceptable incentives. Our estimate translates into the monetary value of 26.4*0.1 QALY gained per 1,000 patients × $22,289 /QALY=

[15] We used a logistic model with pre-integration characteristics of the patient, physician, and area to estimate the propensity score that each physician will integrate and weighed the entire study sample by inverse probability of treatment weights. We truncated the propensity at the 99th percentile to address the unstable weights or for providers with a very low probability of receiving the treatment. Using the weighted sample, we estimated the average treatment effect from our main model.

[16] Using our main model, we examined the hypothetical scenario of all patients who did not receive deep sedation receiving them and predicted the probability of individual patients having adverse outcomes. Based on the predicted probability, we simulated the realized binary outcome of the patient having an adverse outcome, from which we calculated the population-level rates of adverse outcomes per 1,000 colonoscopies.

\$588.4. Thus, paying up to \$588.4 more than the current amount per colonoscopy to prevent the reduction of deep sedation use among integrating physicians can be cost-effective.

The estimated incentive size for providing deep sedation for a colonoscopy is more than half of the current average that Medicare pays to an HOPD for a colonoscopy (\$917). This estimate critically depends on the assumption that providing financial incentives can alter physicians' or practices' behavior in a way that more anesthesiologists would be available to provide deep sedation for colonoscopies. Yet, our estimate can provide an upper bound for the financial incentives that could be used for payment policies. As noted earlier, the current Medicare payment for anesthesiologists is based on their time spent in the operating room, not the procedures for which the anesthesia was performed. Thus, adjusting the price of anesthesia by procedure type might not be directly implementable for Medicare patients. Still, there are other ways to incentivize the providers to adopt more recommended practices, for example, through value-based payment. The estimated incentive size for providing deep sedation, i.e., more than half of the current average that Medicare pays to an HOPD for a colonoscopy, can be significant for the practices and thus, is likely to alter the current physician behavior while remaining beneficial from a societal perspective.

More broadly, our results speak to the recent discussion around the innovative healthcare delivery and financing policies designed to encourage coordination among care providers. For example, Medicare Access and CHIP Reauthorization Act (MACRA) of 2015, which revised physician payment, creates a potential pathway for physicians to earn substantial bonuses for participating in alternative payment models favoring large organizations. Other provider payment reforms such as bundled payment programs or the Federal 340B drug discount program all provide direct or indirect incentives for consolidations among providers in different production segments. Our results provide a cautionary message that when physicians financially integrate in response to these policies that use financial incentives, it does not guarantee that integrated practices will achieve superior patient outcomes. To achieve superior patient outcomes, there should be additional measures to (a) monitor the post-integration physician behavior and quality, and (b) align post-integration financial incentives. For example, CMS could require mandatory reporting of quality measures that are likely to be affected when practices integrate as a part of pay for performance schemes, which would enable them to closely monitor if there are any drastic changes in the integrated organizations' delivery of care. The CMS can also implement payment policies that further promote the provision of high-value care process for the integrated practices to incentivize physicians.

11. Summary of Main Findings and Limitations

Table 8 summarizes our findings, where favorable changes are indicated in blue, and unfavorable changes are indicated in red. Overall, our findings provide evidence that vertical integration adversely affects the quality of care while increasing the healthcare spending of the delivery system by altering physician behaviors. First, the reduction in some value-added care processes (e.g., deep sedation) leads to an increase in some adverse patient outcomes. We further find that this reduction in the value-added care process is, in turn, driven by subsequent constraints in anesthesiologist availability and the inherent financial incentive structure. Second, although there is a sign of an increase in some aspects of the operational efficiency (e.g., throughput), the effect does not result in positive changes in quality or spending. Put together, our results suggest that vertical integration has negative consequences on some important dimensions of healthcare delivery, and hence, requires careful consideration by policymakers.

To assist policymakers, we perform cost-effectiveness analyses on some counterfactual policies using the findings from our mediation analyses. Our results suggest a few ways through which policymakers can effectively adjust the current incentive structure, and thereby mitigate the negative consequences of vertical integration. For example, we find that incentivizing the providers to adopt more recommended practices such as deep sedation is one cost-effective policy lever that can be followed by policymakers. This requires imposing additional regulations to Medicare Access and CHIP Reauthorization Act (MACRA) of 2015 and/or the Federal 340B drug discount program, each of which provides direct or indirect incentives for consolidations among providers through specific payment methods. Policymakers might also require mandatory reporting of extensive quality measures for outpatient practices, which can improve the ability to monitor quality ensure quality does not degrade after integration.

Our study has a number of limitations. First, it focuses on a specific specialty (gastroenterology) and population (FFS Medicare). The findings may differ for other specialty practices that have different quality measures, provider roles, and characteristics of the disease. The younger population or the Medicare managed care population may also have different responses than the FFS population we studied. Second, there are various limitations from the nature of the data we used as well as our empirical strategy of DID. Although we discuss why the concerns for biases from both data and methods are mitigated (Section 6.2) and conduct various robustness checks (Section 9), these do not entirely eliminate all threats to internal validity. For example, our physician efficiency measures are obtained from two different datasets, one of which is available for shorter observation periods of 2012-2015. Although both datasets present a consistent direction for the effect of integration, the data limitation should be taken into account. For measuring quality outcomes, there are various challenges with identifying variations in coding and billing patterns that one needs to consider. Similarly, measuring integration from our data is imperfect and subject to error. There are multiple forms of integration, and using claims data to infer them is inherently challenging.

Table 8: Summary of the Impact of Vertical Integration

	Outcomes	DID effect		Outcomes	DID effect
Process-related quality	Polypectomy	–	Operational efficiency	Time to follow-up	–
	Incomplete col.	–		Time to surgery	–
	Deep sedation	↓		Colonoscopies/year	–
Outcome-related quality	Perforation	–		Any services/year	↑
	Bleeding	↑		Procedure types/year	↑
	Infection	–		Patients/year	↑
	Minor comp.	↑	Spending	Total	↑
	Interval CRC	–		Physician	↓
				Facility	↑

Having a rich dataset that can identify various nuances of integrated entities apart in the future would be helpful. Future research can also contribute by examining the impact of integration among different physician reimbursement structures, identifying the optimal size of incentives, and also by examining how it affects the quality from patients' perspectives. Given the importance of understanding how recent trends in vertical integration impact the healthcare sector, we expect to see more research in this vein in the near future.

References

Allen JI, Kaushal N (2018) New models of gastroenterology practice. *Clinical Gastroenterology and Hepatology* 16(1):3–6.

Arifoğlu K, Ren H, Tezcan T (2020) Hospital Readmissions Reduction Program does not provide the right incentives: Issues and remedies. *Management Science* .

Baicker K, Levy H (2013) Coordination versus competition in health care reform. *New England Journal of Medicine* 369(9):789–791.

Baker LC, Bundorf MK, Devlin AM, Kessler DP (2018) Hospital ownership of physicians: Hospital versus physician perspectives. *Medical Care Research and Review* 75(1):88–99.

Baker LC, Bundorf MK, Kessler DP (2014) Vertical integration: Hospital ownership of physician practices is associated with higher prices and spending. *Health Affairs* 33(5):756–763.

Barnett ML, Keating NL, Christakis NA, O'Malley AJ, Landon BE (2012) Reasons for choice of referral physician among primary care and specialist physicians. *Journal of General Internal Medicine* 27(5):506–512.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1):289–300.

Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny MP, Hill SR (2016) Cost-effectiveness thresholds: Pros and cons. *Bulletin of the World Health Organization* 94(12):925.

Burns LR, Goldsmith JC, Sen A (2014) Horizontal and vertical integration of physicians: A tale of two tails. *Annual Review of Health Care Management*, 39–117 (Emerald Group Publishing Limited).

Capps C, Dranove D, Ody C (2017) Physician practice consolidation driven by small acquisitions, so antitrust agencies have few tools to intervene. *Health Affairs* 36(9):1556–1563.

Capps C, Dranove D, Ody C (2018) The effect of hospital acquisitions of physician practices on prices and spending. *Journal of Health Economics* 59:139–152.

Carlin CS, Dowd B, Feldman R (2015) Changes in quality of health care delivery after vertical integration. *Health Services Research* 50(4):1043–1068.

Centers for Medicare & Medicaid Services (2016) Medicare provider utilization and payment data: Physician and other supplier .

Chan CW, Farias VF, Escobar GJ (2016) The impact of delays on service times in the intensive care unit. *Management Science* 63(7):2049–2072.

Chan CW, Green LV, Lekwijit S, Lu L, Escobar G (2019) Assessing the impact of service level when customer needs are uncertain: An empirical investigation of hospital step-down units. *Management Science* 65(2):751–775.

Chen C, Savva N (2018) Unintended consequences of hospital regulation: The case of the Hospital Readmissions Reduction Program. *Working Paper* .

Chen S, Rex D (2004) Registered nurse-administered propofol sedation for endoscopy. *Alimentary Pharmacology & Therapeutics* 19(2):147–155.

Chronic Condition Data Warehouse (2014) Chronic condition data warehouse .

Clough JD, Dinan MA, Schulman KA (2017) Trends in hospital-physician integration in medical oncology. *The American Journal of Managed Care* 23(10):624–627.

Cooper GS, Xu F, Barnholtz Sloan JS, Schluchter MD, Koroukian SM (2012) Prevalence and predictors of interval colorectal cancers in Medicare beneficiaries. *Cancer* 118(12):3044–3052.

Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, Zauber AG, de Boer J, Fireman BH, Schottinger JE, et al. (2014) Adenoma detection rate and risk of colorectal cancer and death. *New England Journal of Medicine* 370(14):1298–1306.

Cuellar AE, Gertler PJ (2006) Strategic integration of hospitals and physicians. *Journal of Health Economics* 25(1):1–28.

Desai S, McWilliams JM (2018) Consequences of the 340B drug pricing program. *New England Journal of Medicine* 378(6):539–548.

Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. *Medical Care* 36(1):8–27.

Forlines GL (2017) Trends and drivers of physician-hospital integration: The role of Medicare reimbursement. *Working Paper* .

Gaynor M (2014) Competition policy in health care markets: Navigating the enforcement and policy maze. *Health Affairs* 33(6):1088–1093.

Gaynor M, Moreno-Serra R, Propper C (2013) Death by market power: reform, competition, and patient outcomes in the National Health Service. *American Economic Journal: Economic Policy* 5(4):134–66.

Gillies RR, Shortell SM, Anderson DA, Mitchell JB, Morgan KL (1993) Conceptualizing and measuring integration: Findings from the health systems integration study. *Journal of Healthcare Management* 38(4):467.

Graves N, Halton K, Lairson D (2007) Economics and preventing hospital-acquired infection: Broadening the perspective. *Infection Control & Hospital Epidemiology* 28(2):178–184.

Heese HS (2007) Inventory record inaccuracy, double marginalization, and RFID adoption. *Production and Operations Management* 16(5):542–553.

Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, Berry DA (2013) Meta-analyses of colorectal cancer risk factors. *Cancer causes & control* 24(6):1207–1222.

Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, Zwierko M, Rupinski M, Nowacki MP, Butruk E (2010) Quality indicators for colonoscopy and the risk of interval cancer. *New England Journal of Medicine* 362(19):1795–1803.

KC DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

Khiani VS, Soulos P, Gancayco J, Gross CP (2012) Anesthesiologist involvement in screening colonoscopy: Temporal trends and cost implications in the Medicare population. *Clinical Gastroenterology and Hepatology* 10(1):58–64.

Koch TG, Wendling BW, Wilson NE (2017) How vertical integration affects the quantity and cost of care for Medicare beneficiaries. *Journal of Health Economics* 52:19–32.

Kocher R, Sahni NR (2011) Hospitals' race to employ physicians—the logic behind a money-losing proposition. *New England Journal of Medicine* 364(19):1790–1793.

Konetzka RT, Stuart EA, Werner RM (2018) The effect of integration of hospitals and post-acute care providers on Medicare payment and patient outcomes. *Journal of Health Economics* 61:244–258.

Li X, Li Y, Cai X (2013) Double marginalization and coordination in the supply chain with uncertain supply. *European Journal of Operational Research* 226(2):228–236.

Madison K (2004) Hospital-physician affiliations and patient treatments, expenditures, and outcomes. *Health Services Research* 39(2):257–278.

McGuire TG (2000) Physician agency. *Handbook of health economics*, volume 1, 461–536 (Elsevier).

Medicare Payment Advisory Commission (US) (2017) Report to the Congress, Medicare Payment Policy. *Medicare Payment Advisory Commission* .

Meng L, Batt R, Terwiesch C (2018) The impact of facility layout on service worker behavior: An empirical study of nurses in the emergency department. *Manufacturing & Service Operations Management (forthcoming)* .

Neprash HT, Chernew ME, Hicks AL, Gibson T, McWilliams JM (2015) Association of financial integration between physicians and hospitals with commercial health care prices. *JAMA Internal Medicine* 175(12):1932–1939.

Nikpay SS, Richards MR, Penson D (2018) Hospital-physician consolidation accelerated in the past decade in cardiology, oncology. *Health Affairs* 37(7):1123–1127.

Office of Inspector General (2014) Medicare and beneficiaries could save billions if cms reduces hospital outpatient department payment rates for ambulatory surgical center-approved procedures to ambulatory surgical center payment rates. *Department of Health and Human Services Audit Report* (A-05-12-00020).

Pandya A, Doran T, Zhu J, Walker S, Arntson E, Ryan AM (2018) Modelling the cost-effectiveness of pay-for-performance in primary care in the uk. *BMC medicine* 16(1):135.

Pandya A, Soeteman DI, Gupta A, Kamel H, Mushlin AI, Rosenthal MB (2020) Can pay-for performance incentive levels be determined using a cost-effectiveness framework? *Circulation: Cardiovascular Quality and Outcomes* 13(7):e006492.

Post B, Buchmueller T, Ryan AM (2018) Vertical integration of hospitals and physicians: Economic theory and empirical evidence on spending and quality. *Medical Care Research and Review* 75(4):399–433.

Quantin C, Benzenine E, Hägi M, Auverlot B, Abrahamowicz M, Cottenet J, Fournier E, Binquet C, Compain D, Monnet E, et al. (2012) Estimation of national colorectal cancer incidence using claims databases. *Journal of Cancer Epidemiology* 2012.

Rabeneck L, Paszat LF, Hilsden RJ, Saskin R, Leddin D, Grunfeld E, Wai E, Goldwasser M, Sutradhar R, Stukel TA (2008) Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice. *Gastroenterology* 135(6):1899–1906.

Rex DK, Petrini JL, Baron TH, Chak A, Cohen J, Deal SE, Hoffman B, Jacobson BC, Mergener K, Petersen BT, et al. (2006) Quality indicators for colonoscopy. *The American Journal of Gastroenterology* 101(4):873.

Scott KW, Orav EJ, Cutler DM, Jha AK (2017) Changes in hospital-physician affiliations in US hospitals and their effect on quality of care. *Annals of Internal Medicine* 166(1):1–8.

Shapiro C (2010) The 2010 horizontal merger guidelines: From hedgehog to fox in forty years. *Antitrust LJ* 77:49.

Singer SJ, Kerrissey M, Friedberg M, Phillips R (2018) A comprehensive theory of integration. *Medical Care Research and Review* 1077558718767000.

Song LD, Saghafian S (2019) The spillover effects of hospital closures on the efficiency and quality of other hospitals. *Working Paper* .

Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES (2010) Regional variations in diagnostic practices. *New England Journal of Medicine* 363(1):45–53.

Staats BR, Dai H, Hofmann D, Milkman KL (2017) Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science* 63(5):1563–1585.

Tan TF, Netessine S (2019) When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science* 65(8):3495–3517.

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014) Mediation: R package for causal mediation analysis. *Journal of Statistical Software* .

United States Government Accountability Office (2015) Increasing hospital-physician consolidation highlights need for payment reform. *Report to Congressional Requesters* (GAO-16-189).

US Preventive Services Task Force (2008) Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 149(9):627.

Vlachy J, Ayer T, Ayvaci M, Raghunathan S (2017) The business of healthcare: Physician integration in bundled payments. Technical report, Working paper.

Wagner A (2016) Effect of physician-hospital financial integration on health outcomes and spending. *Working Paper* .

Wang J, Zhou YP (2018) Impact of queue configuration on service time: Evidence from a supermarket. *Management Science* 64(7):3055–3075.

Warren JL, Klabunde CN, Mariotto AB, Meekins A, Topor M, Brown ML, Ransohoff DF (2009) Adverse events after outpatient colonoscopy in the medicare population. *Annals of Internal Medicine* 150(12):849–857.