HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

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

Faculty Research Working Paper Series

Soroush Saghafian
Harvard Kennedy School

Linda D. Song
University College London

Joseph P. Newhouse
Harvard Kennedy School, Harvard Medical School, Harvard T.H. Chan School of Public Health

Mary Beth Landrum
Harvard Medical School

John Hsu
Massachusetts General Hospital

September 2020 (Updated November 2022)

RWP20-031

Submitted to
manuscript

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

Soroush Saghafian[1], Lina D. Song[2],
Joseph P. Newhouse[2,3,4,5], Mary Beth Landrum[3], John Hsu[6]
[1] Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA
[2] School of Management, University College London, London E14 5AA, UK
[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 a 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, we examine changes in various measures of care delivery. We find that physicians significantly alter their care process (e.g., in using anesthesia with deep sedation) after they vertically integrate, and there is a substantial increase in patients' post-procedure complications. We further provide evidence that the financial incentive structure of the integrated practices is the main reason for the changes in physician behavior, since it discourages the integrated practices from allocating expensive resources to relatively unprofitable proce-dures. We also find that although integration improves operational efficiency (e.g., measured by physicians' throughput), it negatively affects quality and 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, and highlight some useful policy levers.

Key words: Vertical Integration; Healthcare Operations Management; Healthcare Quality; Provider Payment
History: Version: Nov., 2022

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 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).

Integration involves concurrent changes in multiple dimensions of care delivery, including structure, processes, and people (Singer et al. 2018). Structural integration includes the changes in physical, financial, or legal ties among the providers. Better structural integration can result in the improvement in the care processes and collaboration among providers, which can result in better outcomes, but the relationships among these are difficult to disentangle. 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 one aspect of hospital-physician integration (e.g., changes in the behavioral responses of the physicians that might in turn be induced by forces such as a new financial incentive structure) affects the process, 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 change their care processes (measured by a reduction in deep sedation use) and increase their operational efficiency (measured by throughput). In addition to the changes in the care process, integration adversely affect some patient outcomes. We further provide evidence that such 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 negatively affecting 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 unintended consequences on the healthcare delivery system through both reducing quality and increasing spending.

1.1. Policy Context and Setting

Our study setting is the fee-for-service (FFS) Medicare program, which 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, Medicare reimburses the same procedure in different settings differently: in particular, a procedure that occurs in a hospital outpatient department (HOPD) is paid more than if it took place in a physician's office or an ambulatory surgery center (ASC). 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 an extra amount of money (by about half again as much as the current price for deep sedation) to encourage the use of deep sedation and prevent from changes in the process of care post-integration 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% random sample of all fee-for-service (FFS) Medicare patients between 2008-2015. These data include 2.6 million observations of patient visits provided by 5,488 physicians. We combine the

[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 data with multiple other data sources to examine quality, efficiency, and expenditure outcomes. 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 a specialty area with large outpatient procedure volumes 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 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 clinical 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 variety 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.

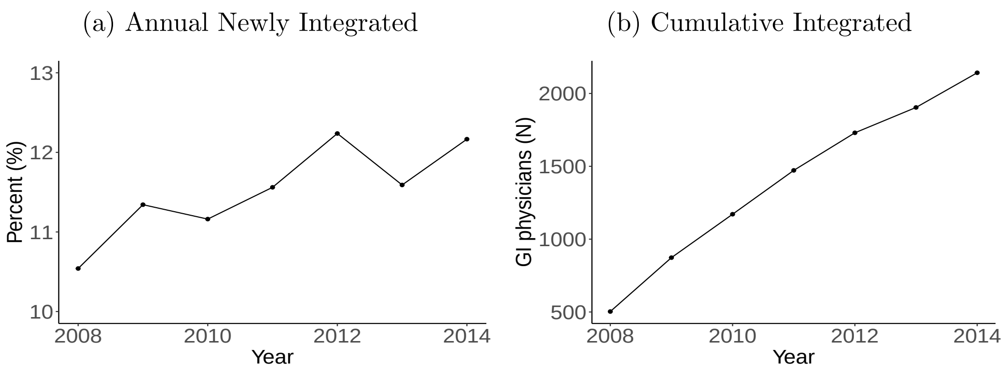
1.3. Main Findings and Contributions

We find that vertical integration negatively affects the quality of care. After physicians integrate, they alter their care delivery process. In particular, our results show that physicians that integrate reduce use of deep sedation, resulting in about 3.7 fewer patients receiving deep sedation per 100

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

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



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.

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 other complications such as cardiac or nonserious GI symptoms (5.0 and 3.3 per 1,000 colonoscopies, respectively). These effects remain directionally the same even after adjusting for changes in patient composition and market characteristics. Through mediation analyses, we find that the reduction in the use of deep sedation at least partially explains the increase in adverse outcomes. Moreover, such changes are driven mainly by hospitals no longer allocating expensive anesthesiologists to relatively unprofitable colonoscopy procedures. In addition, integration increases physicians' throughput and elevates reimbursement per procedure. Notably, we find that integration results in physicians being reimbursed about $127 more per colonoscopy procedure, which is equivalent to an increase of 48% in mean per procedure spending of independent physicians.

Taken together, our results indicate that, despite an increase in spending and throughput, vertical integration does not improve the quality of care. Rather, the shift in the incentive structures of the organization as a result of integration can generate unintended negative consequences in

[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.

both quality and spending. Our cost-effectiveness analyses reveal that increasing the colonoscopy reimbursement rate (by about 150%) to avoid a reduction in the use of deep sedation can be viewed as a cost-effective lever to prevent the unintended negative consequences of integration.

2. Background: Theories of Vertical Integration

Vertical integration refers to the common ownership of two or more stages of production (or distribution) that are initially separate. It is common in healthcare, where primary care physicians refer patients to 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 the organization to achieve any clinical integration, and it could 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 no effect or a negative one. Because the two theories are not mutually exclusive, the pivotal question centers on their respective magnitudes.

Finally, we note that, if integration changes any fee that an individual physician receives, there are 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, Saghafian et al. 2022). Within the context of vertical integration, most studies focus on its 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. The impacts of better aligning hospital capabilities and patient needs, and the policy levers that can yield improvements, have also been discussed in the literature (Saghafian and Hopp 2019, 2020). 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, Atkinson and Saghafian 2019). 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). A vast literature within the operations management studies improving decisions made on routing patients (see, e.g., Saghafian et al. 2019)) or treatments (see, e.g., Boloori et al. 2020). 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. That 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 (see, e.g., 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 healthcare 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 hospital's or the physician's 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.[4]

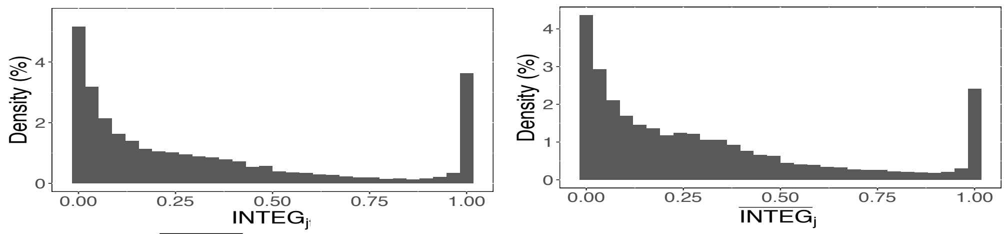
Previous studies of integration that have examined other types of physicians, such as primary care physicians (PCPs) or cardiologists, and have dichotomized integration status into either independent or integrated (Neprash et al. 2015, Desai and McWilliams 2018, Clough et al. 2017). Dichotomization, 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

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

[4] For example, in addition to re-running our analyses by varying thresholds, 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 8.3 for more details).

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.

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. Thus, we separately examine the effect "full integration" later (see Section 8.3).

4.3. Study Population and Comparison Groups

Our patient population is 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.[5] We also focused on the patients who have received colonoscopy from GI physicians.[6] The claims for colonoscopy and related diagnoses were extracted using relevant ICD-9 and HCPCS codes (listed in Table EC1).

5. Variable Definitions and Descriptions

5.1. Outcome Variables

We divide our outcome variables into four categories: process measures, post-procedure complications, operational efficiency, and spending. We measured most outcomes 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 years 2012-2015. All other variables were available for the years 2008-2015. Below, we describe each

[5] 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 they 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.

[6] 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)."

of the four categories of our outcome variables separately. A summary of all our outcome variables is presented in Table 2.

Process Measures. We selected two process measures: polypectomy and incomplete colonoscopy, based on the criteria that (a) they are obtainable from the claims data, (b) are widely accepted measures in terms of process of care, and (c) have the potential to be affected by the known variations among GI physicians' practices.

Studies show that adenoma detection rate (i.e., the proportion of patients with at least one colorectal adenoma detected during colonoscopy) is a well-validated metric for how a colonoscopy is delivered (Corley et al. 2014). However, because of the challenges with measuring it directly from claims data, we instead use polypectomy rate, given that polypectomy rate well correlates with ADR (Gohel et al. 2014, Patel et al. 2013). 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 changes in care processes that can reflect the changes in physician behavior. We identified polypectomy rates from claims by the concurrent pathology bills (Warren et al. 2009). Incomplete colonoscopy[7] was selected based on the indicators endorsed by professional societies (Rex et al. 2006). Incomplete colonoscopies can be a proxy for measuring the process of care delivery, especially because they can result in missed lesions, a contributor to the interval cancer (Cooper et al. 2012).

Another key process measure 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). Deep sedation has not formally been endorsed by professional societies as a quality measure, and hence, we only use it as a process measure to identify potential changes in how care is delivered.[8] In particular, using deep sedation can improve patient experience through the fast onset of action, short duration of action, amnestic effects, faster recovery and discharge times, and increased patient satisfaction (Chen and Rex 2004). Yet, whether the propofol sedation results in better post-procedure clinical outcomes such as lower complication rate has been under debate, and there is a wide variation in the use of propofol for outpatient endoscopy (Pace and Borgaonkar 2018,

[7] We obtained incomplete colonoscopies directly from the HCPCS modifier codes 53, 73, or 74 on colonoscopy claims. 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.

[8] We used the presence of an anesthesiologist or a nurse anesthetist to identify anesthesiology involvement (Cooper et al. 2012, Khiani et al. 2012). 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 Measures	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.
Post-Procedure Complications	Perforation	Incidence of a hole in the wall of part of the GI tract.
	Gastrointestinal bleeding	Major and minor bleeding in the gastrointestinal tract.
	Infection	Incidence of bacterial infections after colonoscopy.
	Cardiovascular	Incidence of cardiovascular symptoms after colonoscopy.
	Serious gastrointestinal	Incidence of serious GI symptoms after colonoscopy.
	Nonserious gastrointestinal	Incidence of nonserious GI symptoms after colonoscopy.
	Interval cancer	Incidence of CRC 6 to 36 months after a negative colonoscopy.
Operational Efficiency	Physician efficiency (throughput)	Total colonoscopies, services, unique procedures, or patients/physician/year.
	Time to complete colonoscopy	Time interval between incomplete to next colonoscopy.
	Time to follow-up	Time interval between positive colonoscopy to follow-up colonoscopy.
Spending	Spending per procedure	Total, physician, or facility spending occurred during the service event.
	Spending per physician	Total colonoscopy related spending occurred/physician/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.

Singh et al. 2008, Adams et al. 2017). Thus, we also separately examined whether the changes in using deep sedation (as a process measure) affect the post-procedure clinical outcomes such as complication rates.

Post-Procedure Complications. We next measured the quality of care using widely accepted procedural complications within a defined period after the surgery. These measures are listed in Table 2, and further details about them, codes we used to identify them (e.g., HCPSC and ICD-9 codes), validations from the medical literature and other sources can be found in Section EC1.

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. We also measured two types of waiting times: time from incomplete colonoscopy to the next follow-up colonoscopy, and time from positive colonoscopy to a follow-up colonoscopy.[9] 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 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.,

[9] 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.

treatment intensity) that affects spending. For example, the 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

Table 3 provides a summary of three types of independent variables we used in our models: patient, physician, and market characteristics.

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).[10] 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. We identified the Ambulatory Surgery Center (ASC) status of each physician's practice based on whether the practice submitted any ASC-based claims. Finally, we included the location of each colonoscopy (e.g., HOPD, office, or ASC) as an independent variable in the analysis to determine whether the changes in outcomes for integrated practices are driven by the shift to more HOPD-based procedures, or whether the changes spill over to procedures in other locations.

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 the total number of GI physicians per person by county from AHRF. Finally, we included county

[10] 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
Location of service	Factor, HOPD, Office, or ASC	Medicare 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

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 post-procedure complications. 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 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

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, ≥1 (%)	95.97	95.80	96.69
Process Measures			
Polypectomy (%)	60.15	59.82	61.41
Incomplete colonoscopy (%)	1.66	1.62	2.11
Deep sedation (%)	57.81	41.00	27.62
Post-Procedure Complications			
Perforation (%)	0.12	0.16	0.20
Bleeding (%)	27.24	29.77	26.71
Infection (%)	0.18	0.25	0.35
Cardiovascular (%)	3.16	4.50	4.53
Serious gastrointestinal (%)	5.90	5.99	4.77
Nonserious gastrointestinal (%)	26.5	28.2	25.7
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 after incomplete (days)	27.70	27.27	26.45
Median time to follow-up (days)	180.2	186.3	184.1
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.

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.[11] To perform our DID analysis, we made use of the following model:

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

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

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.

For some of the outcome variables that are measured at the physician level such as throughput and spending, we used the following model:

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

where $\mathbf{X}$ is the vector of aggregated patient characteristics for each physician j in year t.

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 EC1 confirms that there are similar trends in outcomes before the physicians integrate, and hence, the parallel trend assumption is fairly satisfied. 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. However, for outcomes that show significant changes based on our DID analysis such as deep sedation, bleeding, and other 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 EC2). None of the interaction terms is significant, suggesting no significant difference in time trends between the comparison groups prior to integration. We additionally test the parallel trend assumption using an event study specification (for two years pre- and post-integration), which shows that the integration effects appear after the first year of integration and are consistent across the two years (the details can be found in the Appendix).

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 (see, e.g., variables listed in Table 3), 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.

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. This concern is further mitigated, since we adjust for overall comorbidity, chronic conditions, and sociodemographic status. Furthermore, 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 more with others, which may also be correlated with their quality. It is also possible that 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 using matched physician samples and examining the effect of market conditions (see, e.g., Section 8).

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 may 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 8).

7. Results and Discussion

7.1. Average Effects of Integration

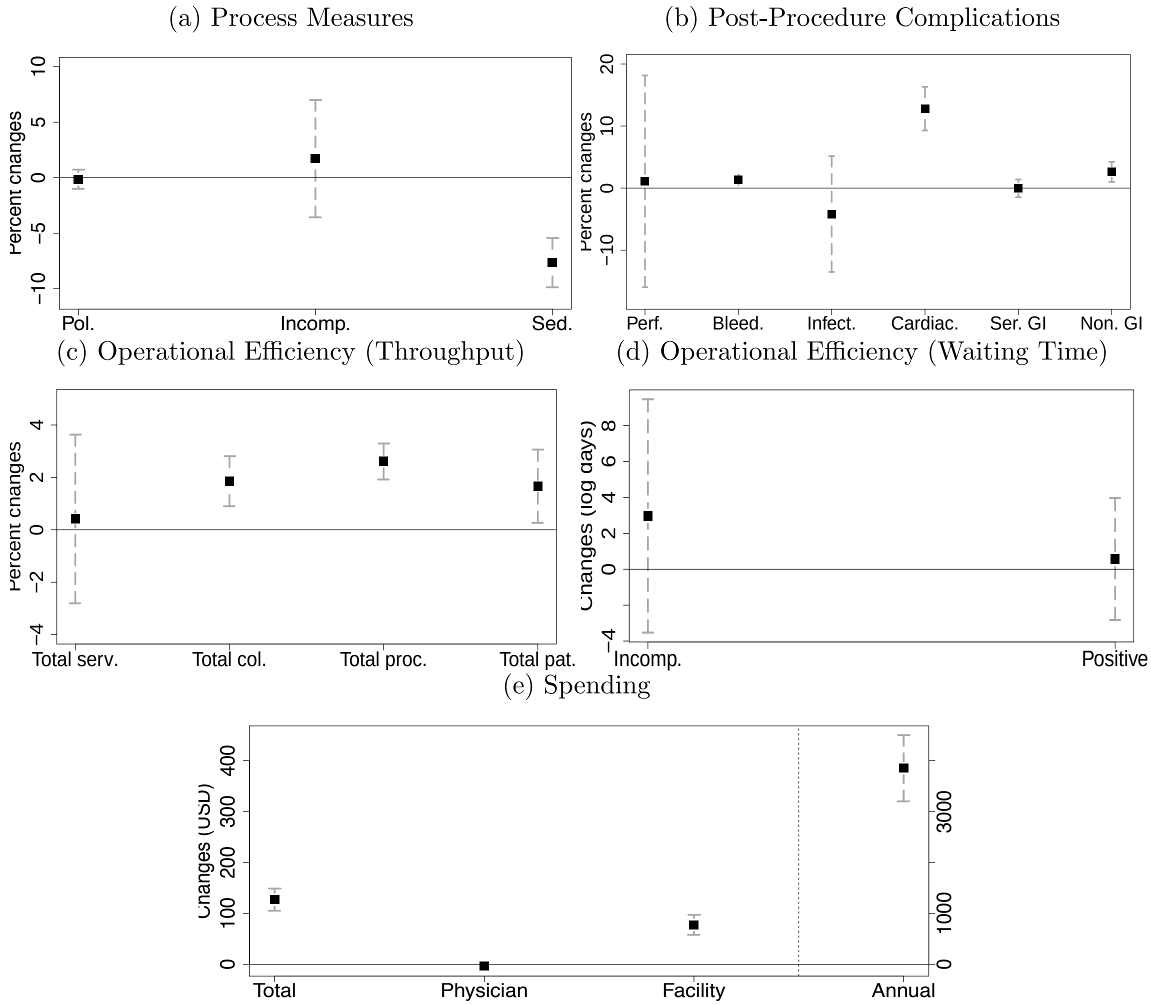
As noted earlier, in our main analysis, we focus on the impact of independent physicians becoming partially integrated, which constitutes the majority of integration cases in our data. Thus, in what follows, we simply label this type of integration as "integration." In later sections, we expand our analyses to other types of integration and provide various robustness checks on our definitions of integration types (see Section 8.3).

Process Measures. Figure 3 and Table 5 show the estimates of our DID coefficients. Full regression results are provided in EC3-EC7. Among the three process measures (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. This effect includes the adjustment for the location of service (HOPD, ASC, etc.).

Post-Procedure Complications. Among various potential complications, patients experience a significant increase in bleeding, cardiac symptoms, and nonserious GI symptoms after colonoscopy when their physicians integrate. Because these complications are relatively common (e.g., the average 30-day incidence of bleeding, cardiac symptoms, and nonserious GI symptoms are 28.6%, 3.94%, and 26.8%, respectively), such increases translate into about 3.8, 5.0, and 3.3 additional bleeding, cardiac symptoms, and nonserious GI symptoms out of 1,000 colonoscopies, respectively.

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, by conducting multiple testing adjustments for all of the 10 process- or outcome-related quality measures that are significant under the assumption of independent hypotheses. We used Bonferroni, the most conservative multiple comparison test (Benjamini and Hochberg 1995), and the family-wise error rates remain significant for three of the four outcomes that were significant under our main model (i.e., deep sedation,

Figure 3: DID Estimates: Average Effects of Integration



Note. Pol. indicates polypectomy. Incomp. indicates incomplete colonoscopy. Sed. indicates deep sedation. Perf. indicates perforation. Bleed. indicates bleeding. Infect. indicates infection. Cardiac. indicates cardiovascular symptoms. Ser. GI indicates serious GI symptoms. Non. GI indicates nonserious GI symptoms. 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.

bleeding, and cardiac symptoms) and non-serious GI symptoms remained marginally significant (see EC15).

Operational Efficiency. Overall, when GI physicians integrate, there are no significant changes in the waiting time variables we measure, either to a follow-up colonoscopy after incomplete or positive colonoscopy (Figure 3). Because these two outcomes are limited to the patients who received two procedures from the same organization, we additionally examined the potential selection by comparing the follow-up rate by providers' integration status. Table EC17 shows that the follow-up rates do not differ significantly based on the integration status.

Table 5: Regression Results

	Dependent variable: process measures and post-procedure complications									
	Pol.	Incomp.	Deep Sed.	Perf.	Bleed.	Infect.	Cardiac.	Seri. GI	Non. GI	Int. CRC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST	−0.00084	0.00028	−0.037***	0.00002	0.0038**	−0.00009	0.005***	−0.0001	0.0033*	−0.00014
	(0.0017)	(0.00044)	(0.0059)	(0.00012)	(0.0012)	(0.00010)	(0.0007)	(0.0020)	(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	2,442,582	2,442,582	1,551,886
R^2	0.097	0.032	0.529	0.013	0.127	0.022	0.068	0.120	0.102	0.019
Adj. R^2	0.092	0.026	0.527	0.007	0.122	0.016	0.068	0.116	0.098	0.012
Res. Std. Err.	0.467	0.125	0.344	0.037	0.424	0.047	0.187	0.420	0.316	0.056

	Dependent variable: operational efficiency and spending									
	Total Serv.	Total Col.	Total Proc.	Total Pat.	Incomp. Time	Follow. Time	Total Spend.	Phy. Spend.	Fac. Spend.	Ann. Spend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST	5.146	3.241**	1.059**	7.596*	0.152	0.078	127.0***	−3.5***	77.5***	3,851.0***
	(20.460)	(0.852)	(0.142)	(2.260)	(0.135)	(0.087)	(11.1)	(0.54)	(10.0)	(331.9)
Obs.	24,025	66,908	34,398	33,337	12,968	16,124	2,442,582	2,442,582	976,456	67,362
R^2	0.924	0.871	0.922	0.947	0.682	0.641	0.359	0.263	0.365	0.816
Adj. R^2	0.871	0.845	0.889	0.924	0.548	0.504	0.355	0.259	0.357	0.778
Res. Std. Err.	721.685	52.363	5.739	78.940	1.114	1.186	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; Cardiac. indicates cardiovascular symptoms; Seri. GI indicates serious gastrointestinal symptoms; Non. GI indicates nonserious gastrointestinal symptoms; Int. CRC indicates interval CRC; Serv. indicates services; Col. indicates colonoscopy; Proc. indicates procedures; Pat. indicates patients; Follow. indicates follow-up; 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.

There are noticeable changes in the throughput measures: the GI physicians significantly increase three out of four measures of throughput (colonoscopy services, procedure types, and patients) after integration. Timely follow-up after incomplete or positive colonoscopy is an important indicator of quality (Winawer et al. 2006). Thus, the fact that follow-up interval does not improve despite the physicians increasing their throughput suggests that the increase in throughput is less likely to be an indication of improved care quality, but a response to changes in financial incentives (e.g., greater payment for colonoscopies performed at integrated settings). We further examine the mechanisms behind the increase in throughput in Section 7.4.

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 reducing the provision of cheaper polypectomy methods and increasing 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 specific types of polypectomy methods[12]) before and after a physician integrated, which suggests that any changes in the procedure mix did not drive up the spending (see EC8). 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 other post-procedure complications. How can policymakers avoid these unintended negative consequences of integration? To answer this question, we next examine the mechanisms behind our results.

7.2. Mechanisms for Changes in Post-Procedure Complications

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 in the medical literature. Yet, we examine the possibility that the reduction in deep sedation among integrated providers may have resulted in increased patients' complications. To examine this, we first included throughput as an independent variable and re-examined the effect of integration on patients' outcomes using the following model:

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

where $THRU_{jt}$ is the colonoscopy throughput per year for physician j in year t, and all variables are as previously defined. For examining the role of deep sedation, we use the following model:

[12] 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).

$$Y_{ijt} = \alpha POST_{jt} + \delta SED_{ijt} + \beta \mathbf{X_{ijt}} + \gamma \mathbf{Z_{it}} + PHYSICIAN_j + MARKET_i + YEAR_t + \epsilon_{ijt}, \quad (5)$$

where SED_{ijt} is the binary variable for whether deep sedation was accompanied.

Our results presented in Figure 4 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 (and not changes in throughput, e.g., a speed-up behavior) can at least partially explain an increase in some of the adverse patient health outcomes.

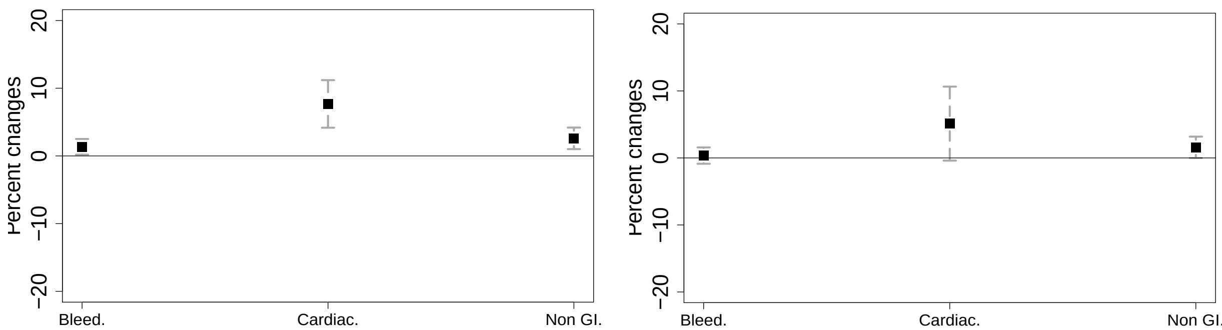
To further provide support for our claims, we implemented simple mediation models using the R package for causal mediation analysis developed by (Tingley et al. 2014). Using a bootstrapping analysis of 10,000 iterations, we estimated the direct and indirect effects of integration via each of the potential mediators (throughput and deep sedation use) on our outcomes of interest (bleeding, cardiac complications, and nonserious GI complications). The results in Table EC9 confirm that integration had a significant effect (i.e., the 95% confidence interval for the size of indirect effect is positive and excludes zero) on deep sedation use, which in turn had a significant impact on both bleeding and other complications. When we examined the mediation effect of throughput, however, we observed that throughput does not mediate the effect of integration on these adverse outcomes.

7.3. Mechanisms for Changes in Process Measures

Given our finding that the reduction in deep sedation use post-integration partially explains an increase in adverse patient outcomes, we next take a closer look at the potential drivers of the changes in deep sedation itself. First, we examine whether the changes in physician behavior in using deep sedation are driven by the changes within the integrated practice, or if the changes are consistent across different practices. To examine this, we included an interaction term between the location of each colonoscopy (HOPD, OFFICE, or ASC) and integration to see if office procedures have a relative reduction in the use of deep sedation if the physician integrates. We find that the interaction term does not turn out to be significant, suggesting that the integration effect is driven by the changes that occur at the integrated practice, but the change does not spill over to other locations in which the integrated physicians practice (see Table EC10). In other words, this suggests that the reduction in deep sedation use is likely driven by the factors associated with the practice, as opposed to the factors associated with the physician.

Figure 4: DID Estimates: Mediation Analyses

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



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.

Based on such findings, we next examine what factors of the HOPD setting drive the reduction in deep sedation use. 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.[13] We found that, on average, a typical GI physician 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

[13] 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 physician and an anesthesiologist or a nurse anesthetist were considered to have worked together if they treated the same patient on the same date.

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 sedations for colonoscopy per anesthesiologist per year, adjusting for 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 Table EC11 for more details). Thus, these results suggest that while integrated practices do not necessarily reduce the number of affiliated anesthesiologists, they shift their use of anesthesiologists to services other than colonoscopies. But why do they do so?

At least two explanations are possible. First, providers may shift the allocation of anesthesiologists because of the 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 are higher than that for colonoscopy. This, however, is unlikely because Medicare pays anesthesiologists as a function of their time spent in the operating room. Thus, there is little financial 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 fixed, and there are other specialty procedures that are more profitable than colonoscopy (e.g., orthopedics or pain management), the marginal financial benefit of providing deep sedation for colonoscopy procedures through anesthesiologist may be low. In that case, organizations may allocate anesthesiologists to the services that generate greater overall revenue. To test this possibility we examined the heterogeneity in the reduction of anesthesia among the practices with and without other specialties and found 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 Table 6 for more details). This result implies that the 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, however, have access to data to directly establish a casual relationship in this regard and so leave it to future research to examine this phenomenon more rigorously. However, our conversations with gastroenterologists also confirm 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 9.

7.4. Mechanisms for Changes in Operational Efficiency

Our main result shows that after integration, physicians increase their operational efficiency (throughput). We hypothesize that the increase in efficiency may have been driven by factors such as the changes in the use of deep sedation or changes in the financial incentives. Although deep sedation is known to reduce the recovery time, it may increase the preparation time, and there is limited evidence on whether administering deep sedation for colonoscopy reduces the entire service time. By including deep sedation use in our main model, we can examine whether the changes in throughput are driven by the changes in the use of deep sedation. Because throughput is defined at the physician level, we calculated a per physician aggregate measure of deep sedation use as the total number of colonoscopies that accompanied deep sedation out of the total number of colonoscopies provided by the physician in a given year.

Consistent with our main analysis, we find that integration increases throughput even after controlling for deep sedation use. Since deep sedation use is also associated with an increase in throughput (see Table EC12), this suggests that, if integration did not result in a reduction in deep sedation, there could have been an even greater increase in throughput. We were not able to directly verify that an increase in financial incentive increases the throughput, given that the price differentials were administratively set and consistent across all providers. Yet, the fact that throughput increases controlling for deep sedation use suggests that the increase in throughput is likely driven by the increase in financial incentives rather than a reduction in deep sedation use. In particular, our finding implies that increasing payment for providing deep sedation can not only result in improvement in quality, but also an increase in operational efficiency. We further examine the effect size in our counterfactual analysis.

8. Robustness Checks

To test the 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, and inverse probability weighting as well as coarsened exact matching 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 EC16). The specific methods and results are provided in the Online Appendix. However, we next briefly

Table 6: Heterogeneity in the Reduction of Anesthesia by Single vs. Multispecialty Practices

	Dependent variable:	
	sed	
	(1) Single-specialty	(2) Multi-specialty
post	$-0.019^{\dagger}$	-0.037^{**}
	(0.010)	(0.007)
Observations	1,805,880	636,702
R^2	0.527	0.590
Adjusted R^2	0.524	0.586
Residual Std. Error	0.345 (df = 1795421)	0.322 (df = 630123)
Note:	†p<0.1; *p<0.05; **p<0.01; ***p<0.001	

discuss each of our main robustness checks. Overall, the results in this section give us confidence that our results are fairly robust and are not sensitive to our assumptions and model specifications.

8.1. 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:

$$Y_{jt} = \alpha INTEG_{jt} + \beta \mathbf{X_{jt}} + \gamma \mathbf{Z_{jt}} + PHYSICIAN_j + YEAR_t + \epsilon_{jt}, \qquad (6)$$

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 same as our main model.

Figure EC2 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.[14] Of note,

[14] 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

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.

8.2. Physician Behavior

Pre-Integration Characteristics. We adjusted for the baseline differences in physicians and patients using inverse probability weighting (IPW).[15] Importantly, to address the possibility that physicians with a different propensity for deep sedation adoption at baseline may make use of deep sedation differently after integration, we included physician's pre-treatment adoption of deep sedation as one of the matching criteria. We also evaluated the results with the matched sample using the coarsened exact matching method. Our results using both IPW and coarsened exact matching are fairly consistent with our main findings (see Table 7).

Coding and Gaming. Another important consideration for 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 effect of integration on two outcomes, bleeding and minor complications. Next, we examined limiting the definition of bleeding to major bleeding only, which is 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 affects our main findings. Further, there is no good reason for physicians to change their coding behavior with respect to these two outcomes, because they are not used for payment or other factors that can create specific incentives.

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.

[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 using R package ipw (van der Wal et al. 2011).

Table 7: Robustness Checks Results

	Deep sedation	Bleeding	Cardiac	Nonserious GI
Behavior				
IPW	−0.037*** (0.005)	0.004** (0.001)	0.010*** (0.001)	0.003* (0.001)
Coarsened exact matching	−0.036** (0.006)	0.008* (0.004)	0.011*** (0.001)	0.004** (0.001)
Major bleeding	NA‡	0.004† (0.0017)	NA‡	NA‡
Exclude retiring physicians	−0.037*** (0.006)	0.004* (0.002)	0.003* (0.001)	0.003* (0.001)
Threshold				
Cutoffs at 1%, 99%	−0.015* (0.005)	0.004** (0.001)	0.001† (0.001)	0.014*** (0.001)
Cutoffs at 5%, 95%	−0.041** (0.008)	0.001† (0.001)	0.002* (0.001)	0.013*** (0.001)
Cutoffs at 15%, 85%	−0.039*** (0.007)	0.003** (0.001)	0.003* (0.001)	0.010*** (0.001)
Binary	−0.039*** (0.0057)	0.005* (0.0016)	0.003** (0.001)	0.0041** (0.0011)
Continuous	−0.139*** (0.020)	0.011 (0.007)	0.004* (0.002)	0.001† (0.001)
First instance	−0.242*** (0.026)	0.063 (0.008)	0.006* (0.003)	0.002* (0.001)
Competition				
Low	−0.031* (0.009)	0.005* (0.002)	0.002* (0.001)	0.004* (0.001)
High	−0.041*** (0.005)	0.004† (0.002)	0.002† (0.001)	0.003† (0.002)

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.

Reduced Work Time. 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 care differently from others. We identified a total of 1,131 GI physicians (6.9% of the GI physicians in our data) who are likely to be on a path to retirement[16] 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).

8.3. Defining Integration

Defining 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 used 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. Next, we used a 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

[16] We defined the GI physicians who are highly likely to retire as the physicians who have submitted any claims for at least one year, but submitted no claims (including inpatient claims) for all subsequent years.

main analysis, likely because the majority of integration cases in our data are related to partial integration, as we noted earlier. We also directly used the continuous integration measure defined in Equation (1) without imposing any threshold. This also showed consistent results with our main findings. Finally, we defined an integration event as the first instance of the use of an HOPD-based reimbursement for a physician who never had an HOPD-based reimbursement before in the dataset, and defined the procedures before the date as pre-integration, and the ones after the event as post-integration. The main effects were still consistent under such definition.

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 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 Table EC14). 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 result in 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 are insufficient to rigorously test these and other possibilities, and we leave it to future research to shed further light on these issues.

8.4. 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. Policy Implications

Overall, our results overall paint a negative picture of vertical integration, showing that it decreases some important aspects of quality while increasing spending. Our evidence also suggests that the negative impact of vertical integration is driven by the physicians' responses to 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, if the providers did not make behavioral changes after integration. If all GI physicians who did not provide deep sedation had done so, about 26 (SE: 1.2) bleeding, 29 (SE: 1.5) cardiac symptoms, and 12 (SE: 1.0) and nonserious GI symptoms per 1,000 colonoscopies would have been averted.[17] 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 analysis (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 one could prevent the reduction of deep sedation use after integration by increasing financial incentives. Assuming 0.1 QALY as the maximum amount of utility loss associated with the post-colonoscopy adverse outcomes (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= $588.4. Thus, adding extra money up to $588.4 to the current reimbursement amount for each colonoscopy to encourage the use of deep sedation among integrating physicians can be cost-effective.

In addition to the impact on quality, we also found that the increase in deep sedation can also increase the operational efficiency (throughput) of physicians. Thus, we estimated the effect of an increase in deep sedation use on the throughput using a similar counterfactual analysis. If integrated physicians do not reduce deep sedation use, each physician will have an additional 2.3 (SE:0.2)

[17] 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. Bootstrapping method was used to estimate the standard error.

increase in throughput. It is difficult to convert this number directly into monetary value, but given the current concerns on the supply of GI physicians that deter patients' access to colonoscopy, an increase in throughput as a result of integration suggests the social impact could be positive.

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 an improvement in process or outcome measures is possible, i.e., more anesthesiologists being available to provide deep sedation. 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 to provide 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. Finally, it is also important to monitor whether such changes in financial incentive creates new unintended consequence by moving away the anesthesia use from other services for which anesthesia is also valuable. To do so, it will be important to have a comprehensive approach in evaluating the incentive structure of integrated practices.

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, the 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 monitor whether there are any drastic changes in the integrated organizations' delivery of care. Monitoring the quality is also critical if additional money is spent in order to incentivize the integrated practices to improve quality. The CMS can also implement payment policies that further promote the provision of a high-value care process for the integrated practices to incentivize physicians.

In closing this section, we highlight that some of the policies we find useful based on our data and analyses might have unintended consequences, and thus, more research is needed prior to implementing them. For example, a dramatic increase in reimbursement for deep sedation may lead to an inappropriate use of it by providers and others. Similarly, it might increase the number of unnecessary procedures overall, leading to significant costs due to overtreatment. In addition, as we discuss in the next section, our study has significant limitations, and it is important for policymakers to take such limitation into account. Finally, policymakers should note that our findings regarding the use of propofol and its mediating effect on adverse quality outcomes (e.g., perforation, gastrointestinal bleeding, interval cancer, infection incidence, the incidence of cardiovascular symptoms and serious or non- serious GI symptoms after colonoscopy) should only be interpreted in the context of integration (and not as recommendations for the typical medical practice).

10. 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, by altering physicians behavior, vertical integration adversely affects the quality of care while increasing spending.

Our various sensitivity analyses show that our main findings are fairly robust, and are unlikely to be importantly affected by our model specifications or assumptions. However, it is important to note that 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. Thus, caution is needed if one wants to generalize our results. 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 8), 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 only for the shorter observation periods 2012-2015. Nonetheless, 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

Table 8: Summary of the Impact of Vertical Integration

	Outcomes	DID effect		Outcomes	DID effect
Process Measures	Polypectomy	–	Operational efficiency	Time to complete	–
	Incomplete col.	–		Time to follow-up	–
	Deep sedation	↓			
Post-Procedure Complications	Perforation	–		Colonoscopies/year	–
	Bleeding	↑		Any services/year	↑
	Infection	–		Procedure types/year	↑
	Cardiac comp.	↑		Patients/year	↑
	Serious GI	–	Spending	Total	↑
	Nonserious GI	↑		Physician	↓
	Interval CRC	–		Facility	↑

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. Having a rich dataset that can more reliably identify various characteristics of integrated entities apart in the future would be helpful. Finally, while we observe strong evidence behind the mechanism we identify, there might be other hidden changes post-integration that we cannot fully rule out. Thus, other possibilities should not be ignored. We, thus, leave it to future research to run specific controlled experiments or use other ways to further test the validity of our findings. Future research can also contribute by examining the impact of integration among different physician reimbursement structures, identifying the optimal size of incentives, and investigating how vertical integration 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

Adams MA, Prenovost KM, Dominitz JA, Holleman RG, Kerr EA, Krein SL, Saini SD, Rubenstein JH (2017) Predictors of use of monitored anesthesia care for outpatient gastrointestinal endoscopy in a capitated payment system. *Gastroenterology* 153(6):1496–1503.

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* .

Atkinson M, Saghafian S (2019) Who should see the patient? on deviations from preferred patient-provider assignments in hospitals. *Working Paper, Available at SSRN* .

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.

Boloori A, Saghafian S, Chakkera HA, Cook CB (2020) Data-driven management of post-transplant medications: An ambiguous partially observable Markov decision process approach. *Manufacturing & Service Operations Management* 22(5):1066–1087.

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.

Gohel TD, Burke CA, Lankaala P, Podugu A, Kiran RP, Thota PN, Lopez R, Sanaka MR (2014) Polypectomy rate: a surrogate for adenoma detection rate varies by colon segment, gender, and endoscopist. *Clinical Gastroenterology and Hepatology* 12(7):1137–1142.

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).

Pace D, Borgaonkar M (2018) Deep sedation for colonoscopy is unnecessary and wasteful. *CMAJ* 190(6):E153–E154.

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.

Patel NC, Islam RS, Wu Q, Gurudu SR, Ramirez FC, Crowell MD, Faigel DO (2013) Measurement of polypectomy rate by using administrative claims data with validation against the adenoma detection rate. *Gastrointestinal Endoscopy* 77(3):390–394.

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.

Robinson JC, Brown TT, Whaley C, Finlayson E (2015) Association of reference payment for colonoscopy with consumer choices, insurer spending, and procedural complications. *JAMA Internal Medicine* 175(11):1783–1789.

Saghafian S, Hopp WJ (2019) The role of quality transparency in health care: Challenges and potential solutions. *NAM Perspectives. Commentary, National Academy of Medicine, Washington, DC.* .

Saghafian S, Hopp WJ (2020) Can public reporting cure healthcare? the role of quality transparency in improving patient-provider alignment. *Operations Research* 68(1):71–92.

Saghafian S, Kilinc D, Traub SJ (2019) Dynamic assignment of patients to primary and secondary inpatient units: Is patience a virtue? *Working Paper (HKS Working Paper No. RWP17-010), Harvard University, Cambridge, MA.* .

Saghafian S, Song L, Raja A (2022) Towards a more efficient healthcare system: Opportunities and challenges caused by hospital closures amid the COVID-19 pandemic. *Health Care Management Science* 25:187–190.

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.

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

Singh H, Poluha W, Cheang M, Choptain N, Inegbu E, Baron K, Taback SP (2008) Propofol for sedation during colonoscopy. *Cochrane database of systematic reviews* (4).

Song LD, Saghafian S (2019) The spillover effects of hospital closures on the efficiency and quality of other hospitals. *Working Paper (HKS Working Paper No. RWP19-006), Harvard University, Cambridge, MA* .

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.

van der Wal WM, Geskus RB, et al. (2011) ipw: an r package for inverse probability weighting. *J Stat Softw* 43(13):1–23.

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.

Winawer SJ, Zauber AG, Fletcher RH, Stillman JS, O'brien MJ, Levin B, Smith RA, Lieberman DA, Burt RW, Levin TR, et al. (2006) Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the us multi-society task force on colorectal cancer and the american cancer society. *CA: A Cancer Journal for Clinicians* 56(3):143–159.

Appendix

EC1. Measuring Outcomes

Polypectomy rate: We measured the indication of polypectomy during colonoscopy as a proxy for quality. Adenoma detection is a well-established predictor of interval cancer and is one of the most important colonoscopy quality measures, but measuring adenoma detection requires pathology data that is not available at the national level. We instead used a polypectomy, i.e., whether any polyp was removed during a screening colonoscopy, which correlates well with adenoma detection and interval cancer. Polypectomy was identified from claims using HCPCS codes and the existence of pathology bills consistent with previous studies (see Table EC1).

Incomplete colonoscopy: Incomplete colonoscopies can result in missed lesions, which is considered a key contributor to the interval cancer. We identified incomplete colonoscopies using the CPT modifier code 53, 73, or 74 based on previous studies. CPT modifier 53 indicates the surgical codes or medical diagnostic codes when the procedure is discontinued because of extenuating circumstances. CPT modifier 73 indicates discontinued outpatient/hospital ambulatory surgical center (ASC) procedure prior to the administration of anesthesia. CPT modifier 74 indicates a surgical or diagnostic procedure requiring anesthesia was terminated after the induction of anesthesia or after the procedure was started (e.g., the incision made, intubation started, scope inserted) due to extenuating circumstances or circumstances that threatened the well-being of the patient.

Deep sedation use: Because the FDA limits that propofl can be administered by individuals trained in the administration of general anesthesia, another provider (i.e., an anesthesiologist or nurse anesthetist) is usually present during the endoscopic procedure if propofol sedation is used during a colonoscopy. We used an approach by an existing study Khiani et al. (2012) and Liu et al. (2012) to identify anesthesiology involvement with colonoscopy. All of the studies 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. Although this approach does not specifically identify the use of propofol, it is presumed that most anesthesia-assisted procedures would include this agent. Khiani et al. (2012) additionally counted the presence of the specialty code for an anesthesiologist or certified registered nurse anesthetist (CRNA) on the date of colonoscopy, and Liu et al. (2012) included another HCPCS code ("00740" for anesthesia service associated with upper gastrointestinal endoscopies). The agreements in identifying anesthesia accompanied colonoscopies were high across the three studies.

Major complication: We examined the three most common and important complications after colonoscopy: perforation, bleeding, and infection. Perforation of the colon after colonoscopy is a significant adverse outcome, detected during or soon after colonoscopy and can ultimately result in death. Currently, perforation occurs about 1 out of 1,000 to 1,400 colonoscopies. Bleeding and infection are also common complications after colonoscopy. Evidence shows that the complication rate varies by the endoscopist's characteristics and can be an indicator of the quality of care, including the skills of the endoscopist and the bowel preparation. We identified perforation and infection events using ICD-9 codes that are present either on or within 7 days after colonoscopy, and bleeding events as the presence of ICD-9 codes that are up to 7 days after colonoscopy date since the presence of bleeding can also be a reason for a visit rather than the outcome (see Table EC1). We tested the sensitivity of including or excluding the date of colonoscopy.

Other complications: Other complications were measured using the previously validated algorithm to measure clinical complication rates after colonoscopy (Robinson et al. 2015). These complications include the presence of cardiovascular, serious gastrointestinal, and/or nonserious gastrointestinal diagnoses, based on medical literature (Warren et al. 2009). Cardiac complications include arrhythmia, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension, and shock. Serious gastrointestinal complications include perforation, lower gastrointestinal bleeding, and infection. nonserious gastrointestinal complications include paralytic ileus, nausea, vomiting, dehydration, abdominal pain, diverticulitis, and enterocolitis. The lists of corresponding codes are provided in EC1.

Severity of bleeding: We further examined the severity of bleeding to distinguish between major and minor bleeding based on other studies. We identified major bleeding as either 1) indication of hemoperitoneum or 2) indication of inpatient or emergency department stays with diagnosis codes for gastrointestinal or nonspecific hemorrhage or 3) indication of transfusion within seven days after the test date.

Interval cancer: An interval CRC is a CRC that appears after a negative colonoscopy. An interval CRC may occur due to genetic and clinical factors of the patient, as well as inadequate polypectomy or missed lesions. Thus, a higher interval cancer rate may signal poor provider quality, to an extent the patient risk is adjusted (Kaminski et al. 2010). We identified colonoscopies attributable to the interval cancer as the test that precedes 6 to 36 months before the first diagnosis of CRC, consistent with previous definitions. To measure the incidence of interval cancer, we first identified the overall CRC incidence. We used the algorithm by (Quantin et al. 2012) to find the first presence of diagnosis codes for CRC in inpatient, outpatient, and office visits, and the first date of presentation of diagnosis codes for CRC. Next, we calculated the time interval between the first date of diagnosis and all colonoscopy dates and identified the colonoscopies that precede first cancer diagnosis by 6 to 36 months. For patients who had more than one colonoscopy within this period, the assignment of cancer was based on the date of the first complete colonoscopy performed. To check the sensitivity, we removed the cases where the patient does not have any confirmation colonoscopy within six months before the first date of CRC diagnosis to account for the long interval to confirmatory colonoscopy.

EC2. Robustness Checks

Inverse Probability Weighting (IPW). We adjusted for the baseline differences in physicians, especially the pre-integration characteristics, using IPW. In IPW, each observation is weighted to balance the representation of subgroups within the full data set. The weight is the reciprocal of the predicted probability of the observation being in the group that was observed for each observation. We used a logistic model with pre-integration characteristics of patient, physician, and area to estimate the propensity score that each physician will integrate. Using the propensity score, we weighed the entire study sample by inverse probability of treatment weights defined as $z/e + (1-z)/(1-e)$, where z denotes treatment status and e denotes the estimated propensity score. The weight ranged from minimum 1.05 and maximum 17.70. 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. We also examined the parallel trend of the IPW weighted sample.

Patient Experience. In addition to the outcome variables discussed earlier, we examined the effect of integration on patient experience to test whether integration affects other dimensions of care. To measure patient experience, we used the Outpatient Consumer Assessment of Healthcare Providers and Systems (OAS CAHPS) survey data for HOPDs. The OAS CAHPS survey asks patients who made outpatient visits about various aspects of their experience, including cleanliness, communication, and overall recommendation. Because these surveys were initiated in 2018, we were not able to perform a DID analysis similar to what we have done for our main analyses. Instead, we conducted a matched regression of physicians who did and did not integrate and compared their outcomes. Among the HOPD-affiliated physicians, we performed matching based on their patients' age, demographics, and clinical risk, and physician's own characteristics such as geographic regions. We ran a linear regression on the matched sample for the four patient experience outcomes: facilities and staff, communication, overall rating, and recommendation. Overall, we find that integration is associated with better patient experience measures (see Table EC16). The results could suggest that despite a reduction in the actual care processes, there can be a potential positive impact of vertical integration. Yet, the results need to be interpreted with caution, as we were not able to adjust for unobserved differences among physicians.

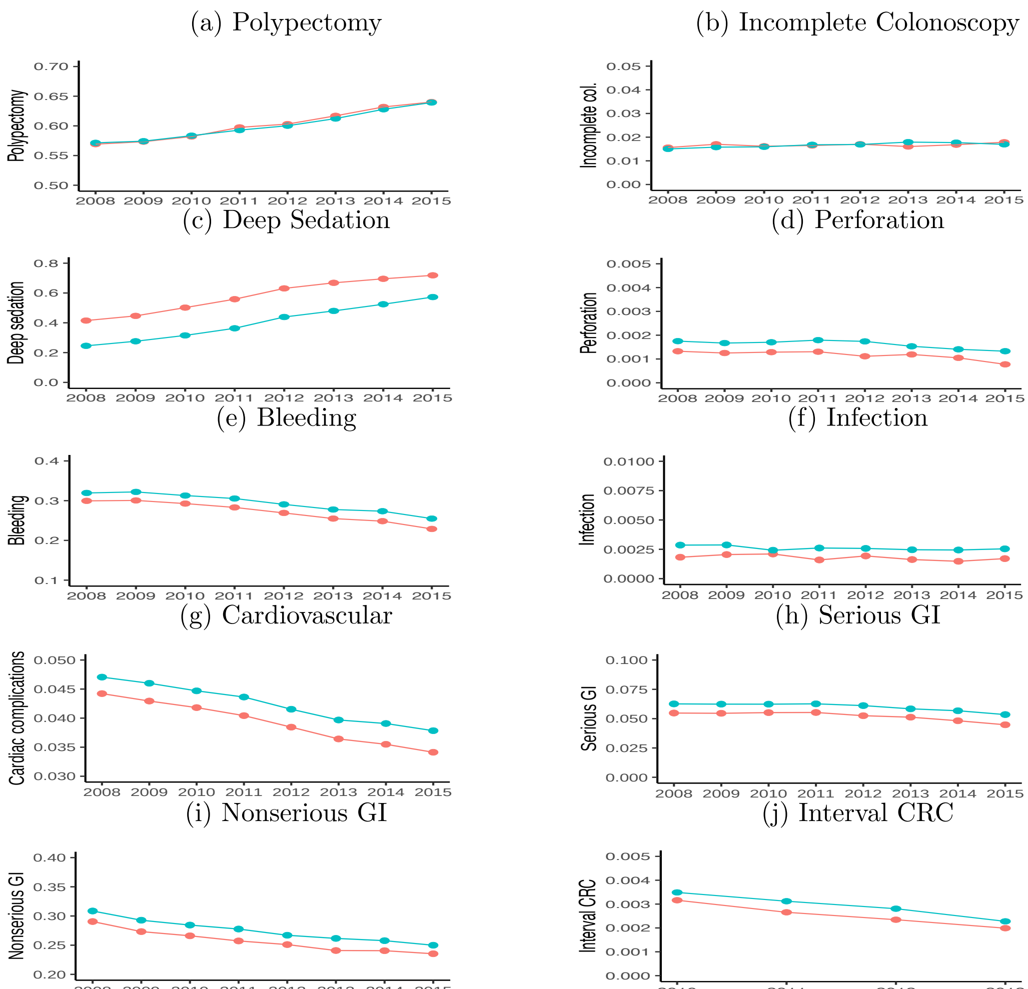
Event Study Specification. We have added the event study specification, presented below and also included in Table EC18 of the manuscript. We examined the effect for the years two years pre and post integration, to ensure that the majority of our sample have the same observation years. We also added this specification and methods in our Appendix. Our results show that the integration effects appear after the first year of integration and are generally consistent across the two years.

$$Y_{ijt} = \alpha_{-2}POST_{j,t-2} + \alpha_{-1}POST_{j,t-1} + \alpha_{1}POST_{j,t+1} + \alpha_{2}POST_{j,t+2} + \beta \mathbf{X_{ijt}} + \gamma \mathbf{Z_{it}} + PHYSICIAN_j + MARKET_i + YEAR_t + \epsilon_{ijt} \quad (7)$$

Controlling for Cardiac-Related Comorbidities. If the underlying patient cardiac risk was higher after physicians integrated, our results might be biased. Therefore, we conducted an additional robustness test that uses models that include cardiac-related comorbidities as an additional covariate. Results can be seen in Table EC19. While the effect size decreases, we still observe that there is a consistent increase in both bleeding and cardiac symptoms post integration.

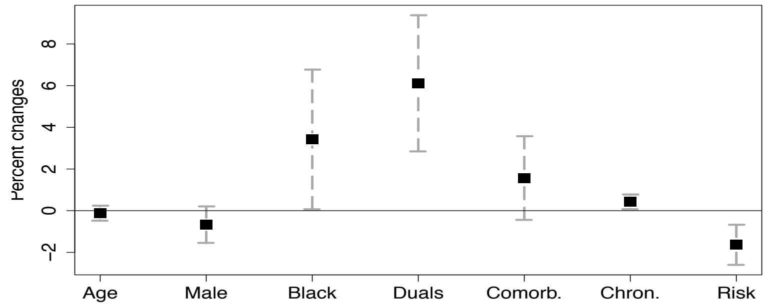
EC3. Figures and Tables

Figure EC1: Trends of Process Measures and Post-Procedure Complications



Note. Red line indicates the trend of independent physicians, and the green line indicates the trend of integrated physicians.

Figure EC2: 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 EC1: List of Codes Used

	Code type	Codes
Colonoscopy	HCPCS	44388, 44389, 44391, 44392, 44393, 44394, 44397, 45378, 45379, 45380, 45381, 45382, 45383, 45384, 45385, 45386, 45391, 45392, G0105, G0121
Polypectomy	HCPCS	44389, 45380, 44392, 45384, 44393, 44394, 45383, 45385, 45379
Incomplete colonoscopy	CPT modifier	53, 73, 74
Deep sedation	HCPCS	00810
Cardiac complications	ICD-9	arrhythmia (ICD-9 codes 427.0-427.4, 427.6-427.9), congestive heart failure (ICD-9 codes 428.0-428.9), cardiac or respiratory arrest (ICD-9 codes 427.5, 799.1, 997.1), syncope, hypotension, or shock (ICD-9 codes 453.29, 458.8-458.9, 639.5, 780.2, 785.50-785.51, 998.0, 995.4)
Serious GI complications	ICD-9	perforation (ICD-9 codes 569.83, 998.2), lower gastrointestinal bleeding (ICD-9 codes 558.9, 578.1, 995.2, 995.89, 998.1-998.13, 286.5, 459, 562.02-562.03, 562.12, 562.13, 569.3, 569.84-569.86, 578.9, 792.1), infection (CPT code 78066 or ICD-9 codes 790.7, 424.9-424.99)
Other GI complications	ICD-9	paralytic ileus (ICD-9 codes 560.1), nausea, vomiting, dehydration (ICD-9 codes 276.5, 536.2, 787.0-02), abdominal pain (ICD-9 codes 789.0), diverticulitis (ICD-9 codes 562.01, 562.03, 562.11, and 562.13), enterocolitis (ICD-9 codes 555-556)
Family history of CRC or polyps	ICD-9	V16.0, V16.8, V16.9, V18.5
History of CRC or polyps	ICD-9	153.0, 153.1, 153.2, 153.3, 153.4, 153.6, 153.7, 153.8, 153.9, 154.0, 154.1, V10.00, V10.05, V10.06
IBD	ICD-9	555.0, 555.1 555.2, 555.9, 556.0, 556.2, 556.3, 556.4, 556.5, 556.6, 556.8, 556.9

Note. CRC indicates colorectal cancer. GI indicates gastrointestinal. IBD indicates inflammatory bowel disease.

Table EC2: Examination of Pre-Treatment Trends

	Dependent variable:			
	Dee sed	Bleed	Cardio	Non GI
	(1)	(2)	(3)	(4)
integ:year2009	−0.001	−0.001	−0.009	0.007
	(0.009)	(0.004)	(0.004)	(0.003)
integ:year2010	−0.005	−0.005	0.003	0.005
	(0.009)	(0.005)	(0.004)	(0.003)
integ:year2011	−0.004	0.001	−0.004	0.003
	(0.008)	(0.005)	(0.004)	(0.003)
Observations	1,380,007	1,380,007	1,380,007	1,380,007
R^2	0.495	0.112	0.124	0.054
Adjusted R^2	0.492	0.108	0.120	0.050
Residual Std. Error (df = 1373997)	0.353	0.341	0.422	0.231
Note:	*p<0.05; **p<0.01; ***p<0.001			

Note. We examined the pre-treatment trends for year 2009-2011 of the providers who have integrated in 2012. The pre-treatment trends of the providers who have integrated in later years (i.e., 2013 or beyond) are similar with the results from the providers who integrated in 2012.

Table EC3: Regression Results: Process Measures

	Dependent variable:		
	Pol	Incomp	Deep Sed
	(1)	(2)	(3)
age	−0.0004*	0.001***	−0.001***
	(0.0001)	(0.00003)	(0.0001)
sex_female	−0.066***	0.004***	0.001
	(0.001)	(0.0003)	(0.001)
race_black	−0.037***	0.002*	−0.011***
	(0.002)	(0.001)	(0.002)
race_other	−0.023***	−0.004***	−0.015***
	(0.003)	(0.001)	(0.002)
race_asian	−0.019***	−0.008***	−0.019***
	(0.002)	(0.001)	(0.003)
race_hispanic	−0.036***	−0.004**	−0.008
	(0.004)	(0.001)	(0.004)
race_native	−0.007	0.0002	−0.007
	(0.007)	(0.002)	(0.006)
esrd	−0.009	0.0003	−0.040***
	(0.005)	(0.002)	(0.006)
dual	0.002	0.008***	−0.017***
	(0.002)	(0.001)	(0.002)
chronic	0.006	−0.0001	0.005
	(0.003)	(0.0004)	(0.009)
comorbid	0.006***	0.002***	−0.007***
	(0.0003)	(0.0001)	(0.001)
rural	0.005*	0.0004	−0.006
	(0.002)	(0.001)	(0.003)
unemp	0.001	−0.0001	0.001
	(0.001)	(0.0001)	(0.0005)
poverty	−0.0001	0.0001	−0.001***
	(0.0001)	(0.00004)	(0.0002)
under_65	−0.0002	0.00004	−0.001**
	(0.0001)	(0.00005)	(0.0002)
gi_supply	−57.883*	5.656	−6.990
	(17.347)	(3.641)	(14.479)
hmo_pen	−0.00002	−0.00000	0.001*
	(0.0001)	(0.00003)	(0.0002)
hhi	0.00000	0.00000	−0.00000*
	(0.00000)	(0.00000)	(0.00000)
plcsrvc_out	0.030***	−0.0002	−0.036*
	(0.004)	(0.001)	(0.012)
plcsrvc_asc	0.032***	−0.0001	0.041**
	(0.004)	(0.001)	(0.012)
post	−0.001	0.0003	−0.037***
	(0.002)	(0.0004)	(0.005)
Observations	2,442,582	2,442,582	2,442,582
R^2	0.097	0.032	0.529
Adjusted R^2	0.092	0.026	0.527
Residual Std. Error (df = 2429138)	0.467	0.125	0.344
Note:			*p<0.05; **p<0.01; ***p<0.001

Table EC4: Regression Results: Post-Procedure Complications

	Dependent variable:						
	Perf	Bleed	Infect	Non GI	Seri GI	Cardiac	Int CRC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
age	0.0001***	0.011***	0.0001***	0.004***	0.009***	0.003***	0.0001**
	(0.00001)	(0.0001)	(0.00001)	(0.0001)	(0.0005)	(0.00004)	(0.00002)
sex_female	0.0005***	0.030***	−0.0003*	0.032***	0.018***	−0.002	−0.0004
	(0.0001)	(0.001)	(0.0001)	(0.001)	(0.001)	(0.001)	(0.0002)
race_black	−0.0004*	0.007***	−0.00002	−0.017***	−0.010*	0.001	0.0005
	(0.0002)	(0.001)	(0.0002)	(0.002)	(0.003)	(0.001)	(0.0004)
race_other	0.0002	−0.004	0.0002	−0.017***	−0.005	−0.008***	0.001
	(0.0003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
race_asian	−0.0004	−0.014***	−0.001	−0.031***	−0.017**	−0.017***	−0.001
	(0.0002)	(0.002)	(0.0005)	(0.003)	(0.003)	(0.001)	(0.0003)
race_hispanic	−0.001	−0.030***	−0.001**	−0.011**	−0.036***	−0.018***	−0.0003
	(0.0004)	(0.003)	(0.0003)	(0.003)	(0.006)	(0.002)	(0.001)
race_native	−0.001	0.040***	0.001	0.006	0.034**	0.012	−0.001
	(0.001)	(0.006)	(0.002)	(0.005)	(0.007)	(0.009)	(0.001)
esrd	−0.0004	0.158***	0.016***	0.091***	0.168***	0.094***	−0.003
	(0.001)	(0.003)	(0.001)	(0.006)	(0.006)	(0.006)	(0.001)
dual	0.0005	0.077***	0.001**	0.048***	0.052***	0.021***	0.0002
	(0.0002)	(0.001)	(0.0003)	(0.002)	(0.001)	(0.001)	(0.0005)
chronic	0.00001	0.006***	−0.0003**	0.0001	0.002	−0.003	−0.0002
	(0.0001)	(0.002)	(0.0001)	(0.001)	(0.003)	(0.003)	(0.0001)
comorb	0.0003***	0.036***	0.001***	0.024***	0.027***	0.019***	0.001***
	(0.00003)	(0.0001)	(0.0001)	(0.001)	(0.001)	(0.0004)	(0.0001)
rural	0.001*	0.013***	0.001*	0.008**	0.013***	0.001	0.0001
	(0.0002)	(0.001)	(0.0002)	(0.002)	(0.002)	(0.001)	(0.0001)
unemp	0.0001	−0.001	−0.00002	0.0001	−0.001	−0.0001	0.0001
	(0.00003)	(0.0003)	(0.00004)	(0.0004)	(0.001)	(0.0001)	(0.0001)
poverty	−0.00003	0.0004***	−0.00001	0.0001	0.0003	−0.0001*	−0.0001*
	(0.00002)	(0.0001)	(0.00003)	(0.0001)	(0.0001)	(0.00005)	(0.00001)
under_65	−0.00000	0.0005***	0.00000	−0.00003	0.0003	0.0003***	−0.00002
	(0.00001)	(0.0001)	(0.00002)	(0.0002)	(0.0001)	(0.00004)	(0.00002)
gi_supply	2.786	−73.752***	−1.894	−53.238***	−73.546**	−32.382**	−2.729
	(1.965)	(10.936)	(1.134)	(9.523)	(17.610)	(6.989)	(1.856)
hmo_pen	−0.00000	−0.0003***	−0.00000	−0.0004**	−0.0004*	0.0001**	−0.00001
	(0.00001)	(0.0001)	(0.00001)	(0.0001)	(0.0001)	(0.00001)	(0.00001)
hhi	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
plcsrvc_out	0.0002	0.034***	0.0002	0.015**	0.021**	0.007***	0.001
	(0.0002)	(0.002)	(0.0002)	(0.003)	(0.006)	(0.001)	(0.001)
plcsrvc_asc	−0.00004	−0.005*	−0.0003	−0.007	0.004	0.001	0.001
	(0.0002)	(0.002)	(0.0002)	(0.004)	(0.007)	(0.001)	(0.001)
post	0.00002	0.004**	−0.0001	0.003*	−0.0001	0.005***	−0.0001
	(0.0001)	(0.001)	(0.0001)	(0.001)	(0.002)	(0.001)	(0.0003)
Observations	2,442,582	2,442,582	2,442,582	2,442,582	2,633,499	2,635,352	1,551,886
R^2	0.013	0.127	0.022	0.102	0.120	0.068	0.019
Adjusted R^2	0.007	0.122	0.016	0.098	0.116	0.068	0.012
Residual Std. Error	0.037 (df = 2429138)	0.424 (df = 2429138)	0.047 (df = 2429138)	0.316 (df = 2429138)	0.420 (df = 2620055)	0.187 (df = 2635325)	0.056 (df = 1539893)

Note: *p<0.05; **p<0.01; ***p<0.001

Table EC5: Regression Results: Operational Efficiency (Throughput)

	Dependent variable:			
	Total Serv	Total Col	Total Proc	Total Pat
	(1)	(2)	(3)	(4)
age	−1.208	−0.574**	−0.005	0.113
	(1.580)	(0.106)	(0.016)	(0.187)
sex_prop_male	24.791	1.567	0.280	2.016
	(37.851)	(2.205)	(0.586)	(5.444)
race_prop_white	38.645	9.530*	1.833	18.884
	(33.966)	(3.165)	(0.741)	(8.158)
prop_dual	−16.123	−8.245*	−0.011	13.370
	(27.567)	(3.386)	(0.539)	(7.572)
comorb	1.206	−3.020***	0.071	−0.370
	(4.632)	(0.461)	(0.047)	(0.720)
chronic	−226.046	8.653	−1.232	−1.088
	(107.807)	(9.534)	(2.024)	(24.242)
prop_rural	39.091	−7.950	−1.275	−29.106
	(46.995)	(5.448)	(1.007)	(11.779)
gi_supply	−1,166,130.000	−33,316.770	−20,937.880	−141,328.400
	(530,952.800)	(20,676.980)	(13,712.540)	(58,161.450)
unemp	−16.618	3.378**	−0.471	3.694
	(13.405)	(0.845)	(0.361)	(2.864)
poverty	3.681	0.074	0.087	1.201
	(8.780)	(0.315)	(0.093)	(0.773)
hmo_pen	−3.565	−2.147***	−0.068	−3.302**
	(2.044)	(0.169)	(0.025)	(0.453)
hhi	0.018	0.003	0.0004	0.010
	(0.019)	(0.001)	(0.0002)	(0.004)
post	5.146	3.241**	1.059**	7.596*
	(20.460)	(0.852)	(0.142)	(2.260)
Observations	24,025	66,908	34,398	33,337
R^2	0.924	0.871	0.922	0.947
Adjusted R^2	0.871	0.845	0.889	0.924
Residual Std. Error	721.685 (df = 14118)	52.363 (df = 55637)	5.739 (df = 24293)	78.940 (df = 23243)
Note:				*p<0.05; **p<0.01; ***p<0.001

Table EC6: Regression Results: Waiting Time

	Dependent variable:	
	Log Incomp Time	Log Follow Time
	(1)	(2)
age	−0.0002	−0.012
	(0.007)	(0.006)
sex_female	0.184*	0.174*
	(0.060)	(0.050)
race_black	0.067	0.151
	(0.104)	(0.070)
race_other	−0.083	−0.057
	(0.152)	(0.154)
race_asian	0.441	0.414
	(0.264)	(0.214)
race_hispanic	−0.003	−0.059
	(0.188)	(0.175)
race_native	−0.080	−0.174
	(0.407)	(0.356)
esrd	0.208	−0.110
	(0.303)	(0.222)
dual	0.061	−0.021
	(0.082)	(0.060)
chronic	0.003	0.143
	(0.145)	(0.130)
comorb	−0.003	−0.037**
	(0.009)	(0.008)
rural	0.018	0.129
	(0.084)	(0.083)
unemp	−0.031	−0.016
	(0.019)	(0.019)
poverty	0.002	−0.0005
	(0.010)	(0.010)
under_65	−0.010	0.001
	(0.009)	(0.008)
gi_supply	−1,184.367	−1,298.640
	(1,123.960)	(1,050.033)
hmo_pen	−0.001	−0.001
	(0.005)	(0.004)
hhi	−0.00002	−0.00004
	(0.00003)	(0.00004)
plcsrvc_out	0.497	0.483
	(0.237)	(0.247)
plcsrvc_asc	0.363	0.360
	(0.215)	(0.226)
post	0.152	0.078
	(0.135)	(0.087)
Observations	12,968	16,124
R^2	0.682	0.641
Adjusted R^2	0.548	0.504
Residual Std. Error	1.114 (df = 9106)	1.186 (df = 11660)
Note:		*p<0.05; **p<0.01; ***p<0.001

Table EC7: Regression Results: Spending

	Dependent variable:		
	Total Spend	Phy Spend	Fac Spend
	(1)	(2)	(3)
age	0.488**	−0.085	2.170***
	(0.112)	(0.060)	(0.165)
sex_female	−4.919***	−3.766***	2.338*
	(0.746)	(0.120)	(0.832)
race_black	−7.013**	−3.250***	−13.223**
	(1.461)	(0.158)	(2.703)
race_other	−2.606	−1.259	−5.312*
	(1.434)	(0.581)	(2.186)
race_asian	4.188	0.561	3.475
	(2.666)	(0.569)	(4.473)
race_hispanic	3.570	−1.738*	−5.340
	(2.385)	(0.694)	(4.824)
race_native	−14.740*	−4.446**	4.032
	(4.681)	(0.912)	(5.615)
esrd	−34.695***	−1.872*	53.221***
	(5.089)	(0.704)	(7.040)
dual	17.901***	−0.442	32.179***
	(1.789)	(0.413)	(1.858)
chronic	19.343	7.774	5.508
	(16.926)	(6.275)	(5.560)
comorb	−1.449	−0.436*	12.012***
	(0.623)	(0.148)	(0.736)
rural	14.636***	−0.619*	7.042*
	(2.034)	(0.257)	(2.767)
unemp	0.470	0.111*	1.772*
	(0.427)	(0.042)	(0.651)
poverty	−0.050	−0.070**	−0.253
	(0.170)	(0.017)	(0.224)
under_65	−0.476*	−0.066**	−0.432
	(0.145)	(0.016)	(0.270)
gi_supply	−83,881.790***	2,318.263	−83,433.070***
	(14,080.990)	(3,046.107)	(13,961.910)
hmo_pen	−0.371**	−0.006	−0.850**
	(0.103)	(0.014)	(0.166)
hhi	0.0004	−0.0002	−0.0001
	(0.001)	(0.0001)	(0.001)
post	126.956***	−3.542***	77.477***
	(11.105)	(0.536)	(10.043)
Observations	2,442,582	2,442,582	976,456
R^2	0.359	0.263	0.365
Adjusted R^2	0.355	0.259	0.357
Residual Std. Error	287.040 (df = 2429138)	60.216 (df = 2429138)	292.382 (df = 963875)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table EC8: Changes in Procedure Mix

	Dependent variable:	
	Prop Biopsy	Prop Pol
	(1)	(2)
post	0.003	0.006
	(0.001)	(0.003)
Observations	34,464	63,502
R^2	0.743	0.526
Adjusted R^2	0.636	0.426
Residual Std. Error	0.045 (df = 24321)	0.183 (df = 52454)
Note:		*p<0.05; **p<0.01; ***p<0.001

Table EC9: Mediation Analysis Results

Mediator	Throughput		Deep sedation	
Outcomes	Bleed	Cardiac	Bleed	Cardiac
Average causal mediation effect	-0.00003 (-0.0002-0.00002)	-0.0003*** (-0.0004-0.00001)	0.001*** (0.00089-0.00004)	0.0013*** (0.0012-0.0000
Average direct effect	0.017*** (0.016-0.02)	0.013*** (0.012-0.01)	0.016*** (0.015-0.02)	0.0090*** (0.0087-0.01)
Total effect	0.017*** (0.016-0.02)	0.012*** (0.011-0.01)	0.017*** (0.016-0.02)	0.010*** (0.009-0.01)
Prop. mediated	-0.0014 (-0.010-0.0000)	-0.023*** (-0.036- -0.01)	0.060*** (0.05-0.07)	0.12*** (0.12-0.13)

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

Table EC10: Regression Results: Interactions with Places of Service

	Dependent variable:
	Deep Sed
post	−0.011
	(0.018)
plcsrvc_out	−0.046*
	(0.017)
plcsrvc_asc	0.044*
	(0.013)
post:plcsrvc_out	0.011
	(0.016)
post:plcsrvc_asc	−0.007
	(0.018)
Observations	2,222,322
R^2	0.567
Adjusted R^2	0.565
Residual Std. Error	0.330 (df = 2209024)
Note:	*p<0.05; **p<0.01; ***p<0.001

Table EC11: Regression Results: (a) Changes in Affiliated Anesthesiologists: External Margin (Left) and (b) Changes in Anesthesiology Volume: Internal Margin (Right)

	Dependent variable:	
	Num. Affiliated Anesthesiologists	
	(1)	(2)
post	0.805^{***}	-0.672^{**}
	(0.121)	(0.149)
Observations	67,362	45,867
R^2	0.775	0.802
Adjusted R^2	0.729	0.750
Residual Std. Error	3.736 (df = 55993)	3.834 (df = 36439)
Note:		*p<0.05; **p<0.01; ***p<0.001

	Dependent variable:	
	Total Anesthesiologists	
	(1)	(2)
post	-2.788^{***}	-1.880^{***}
	(0.218)	(0.295)
Observations	586,368	384,988
R^2	0.235	0.209
Adjusted R^2	0.219	0.188
Residual Std. Error	21.851 (df = 574205)	16.460 (df = 374975)
Note:		*p<0.05; **p<0.01; ***p<0.001

Table EC12: Mechanisms for Changes in Operational Efficiency

	Dependent variable:			
	Total Serv	Total Col	Total Proc	Total Pat
	(1)	(2)	(3)	(4)
post	−27.644	3.014^{*}	0.616	3.749
	(44.850)	(1.227)	(0.284)	(5.201)
sed	51.872	12.071^{***}	−0.501	5.185
	(58.743)	(1.837)	(0.348)	(6.319)
post:sed	60.297	1.274	0.775	6.939
	(52.619)	(1.834)	(0.418)	(5.380)
Observations	24,025	66,908	34,398	33,337
R^2	0.924	0.871	0.922	0.947
Adjusted R^2	0.871	0.845	0.889	0.924
Residual Std. Error	721.597 (df = 14116)	52.310 (df = 55635)	5.739 (df = 24291)	78.929 (df = 23241)
Note:				*p<0.05; **p<0.01; ***p<0.001

Table EC13: Regression Results: (a) IPW Adjusted (Left) and (b) Coarsened Exact Matching (Right)

	Estimate	Cluster s.e.	Pr(>|t|)
Pol	-0.001	0.002	0.633
Incomp	0.0003	0.0004	0.550
Deep Sed	-0.037***	0.005	0.0003
Perf	0.00002	0.0001	0.897
Bleed	0.004**	0.001	0.001
Infect	-0.0001	0.0001	0.413
Cardiac	0.010***	0.001	0.00000
Serioius GI	-0.0001	0.002	0.973
Nonserious GI	0.003*	0.001	0.014
Int CRC	-0.0001	0.0003	0.618

	Estimate	Cluster s.e.	Pr(>|t|)
Pol	-0.009	0.006	0.154
Incomp	0.002	0.001	0.254
Deep Sed	-0.036**	0.006	0.001
Perf	0.00001	0.0002	0.966
Bleed	0.008*	0.004	0.044
Infect	-0.0001	0.0004	0.728
Cardiac	0.011***	0.001	0.0001
Serioius GI	0.002	0.004	0.651
Nonserious GI	0.004**	0.001	0.001
Int CRC	0.001	0.001	0.394

Table EC14: Regression Results: Full Integration

	Dependent variable: process measures and post-procedure complications									
	Pol.	Incomp.	Deep Sed.	Perf.	Bleed.	Infect.	Cardiac.	Seri. GI	Non. GI	Int. CRC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST	−0.0087	0.0011	−0.021	−0.00016	0.0061	0.00044	0.003	−0.0001	0.010	0.00004
	(0.0041)	(0.0013)	(0.011)	(0.00031)	(0.0056)	(0.00075)	(0.004)	(0.002)	(0.001)	(0.001)
Obs.	1,487,344	1,487,344	1,487,344	1,487,344	1,487,344	1,487,344	1,487,344	1,487,344	1,487,344	940,845
R^2	0.096	0.037	0.510	0.019	0.140	0.025	0.108	0.068	0.120	0.025
Adj. R^2	0.089	0.030	0.507	0.011	0.134	0.017	0.101	0.068	0.116	0.015
Res. Std. Err.	0.468	0.126	0.344	0.040	0.425	0.051	0.323	0.187	0.420	0.057

	Dependent variable: operational efficiency and spending									
	Total Serv.	Total Col.	Total Proc.	Total Pat.	Incomp. Time	Follow. Time	Total Spend.	Phy. Spend.	Fac. Spend.	Ann. Spend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST	−111.304**	−9.619**	−4.043**	−28.564*	0.366	0.012	75.354***	−4.367**	24.130**	265.054
	(16.128)	(1.952)	(0.429)	(4.922)	(0.315)	(0.083)	(9.519)	(0.912)	(4.816)	(291.982)
Obs.	18,152	45,502	23,951	23,397	10,009	26,387	1,487,344	1,487,344	939,345	45,867
R^2	0.921	0.877	0.913	0.930	0.682	0.337	0.291	0.147	0.324	0.822
Adj. R^2	0.863	0.845	0.868	0.894	0.543	0.092	0.286	0.140	0.316	0.776
Res. Std. Err.	436.827	51.617	5.857	82.626	1.136	1.133	332.085	58.687	274.980	6,617.175

Note. Pol. indicates polypectomy. Incomp. indicates incomplete colonoscopy. Sed. indicates deep sedation. Perf. indicates perforation. Bleed. indicates bleeding. Infect. indicates infection. Cardiac. indicates cardiovascular symptoms. Ser. GI indicates serious GI symptoms. Non. GI indicates nonserious GI symptoms. 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.

Table EC15: Bonferroni Correction

	Pol	Incomp	Deep Sed	Perf	Bleed	Infect	Non GI	Seri GI	Cardiac	Int CRC
Adj p-val	1.00	1.00	0.003	1.00	0.011	1.00	0.158	1.00	0.00003	1.00

Table EC16: Regression Results: Patient Experience

	Dependent variable:			
	facilities and staff	communication	overall rating	recommendation
	(1)	(2)	(3)	(4)
age	−0.025***	−0.059***	−0.037***	−0.035**
	(0.004)	(0.007)	(0.007)	(0.011)
sex_prop_male	−0.134	0.017	−0.205	−0.199
	(0.070)	(0.117)	(0.125)	(0.195)
race_prop_white	0.400***	0.158	0.092	0.451*
	(0.067)	(0.111)	(0.118)	(0.185)
prop_dual	−0.376***	−0.382**	−1.131***	−1.303***
	(0.084)	(0.141)	(0.150)	(0.235)
comorb	0.012	−0.008	−0.005	−0.065*
	(0.011)	(0.018)	(0.019)	(0.030)
chronic	0.474	1.010*	0.600	0.430
	(0.273)	(0.456)	(0.486)	(0.760)
prop_rural	0.015	0.039	−0.408***	−0.563***
	(0.045)	(0.075)	(0.080)	(0.126)
gi_supply	−501.287*	−859.999*	787.086	3,034.942***
	(234.549)	(391.142)	(416.591)	(652.256)
unemp	−0.028***	−0.101***	0.020	0.009
	(0.007)	(0.011)	(0.012)	(0.018)
poverty	−0.012***	0.013**	−0.005	−0.020**
	(0.002)	(0.004)	(0.004)	(0.007)
hmo_pen	0.001	−0.006***	−0.007***	−0.002
	(0.001)	(0.001)	(0.002)	(0.002)
hhi	0.00002	0.00002	−0.00001	−0.00001
	(0.00001)	(0.00001)	(0.00001)	(0.00002)
post	0.105***	0.029	0.160***	0.150*
	(0.022)	(0.036)	(0.039)	(0.061)
Constant	99.576***	98.538***	96.068***	95.364***
	(0.381)	(0.635)	(0.677)	(1.059)
Observations	5,508	5,508	5,506	5,506
R^2	0.049	0.036	0.026	0.026
Adjusted R^2	0.046	0.034	0.024	0.023
Residual Std. Error	0.746 (df = 5494)	1.244 (df = 5494)	1.325 (df = 5492)	2.075 (df = 5492)
F Statistic	21.573*** (df = 13; 5494)	15.792*** (df = 13; 5494)	11.370*** (df = 13; 5492)	11.160*** (df = 13; 5492)

Note: *p<0.05; **p<0.01; ***p<0.001

Table EC17: Follow-Up Rates for Incomplete and Positive Colonoscopies

% followed up	Independent	Partially integrated	Fully integrated
Incomplete colonoscopy	40.5	39.9	40.8
Positive colonoscopy	23.8	23.1	23.9

Note: Follow-up rates for incomplete colonoscopies are defined as the proportion of incomplete colonoscopies for which there exists at least one other complete colonoscopy within the six month by the same provider, among all incomplete colonoscopies. Follow-up rates for positive colonoscopies are defined as the proportion of positive colonoscopies (e.g., the ones with biopsy for which there exists at least one other colonoscopy within six month by the same provider, among all positive colonoscopies).

Table EC18: Regression Results: Event Study Specification

	Dependent variable:			
	Dee sed	Bleed	Cardiac	Non GI
	(1)	(2)	(3)	(4)
integ:pre2	−0.001	−0.001	−0.009	0.007
	(0.009)	(0.004)	(0.004)	(0.003)
integ:pre1	−0.005	−0.005	0.003	0.005
	(0.009)	(0.005)	(0.004)	(0.003)
integ:post1	−0.041***	0.001*	0.0004***	0.003
	(0.008)	(0.0005)	(0.0001)	(0.003)
integ:post2	−0.032***	0.001***	0.0005***	0.003*
	(0.001)	(0.0001)	(0.0001)	(0.001)
Observations	1,380,007	1,380,007	1,380,007	1,380,007
R^2	0.495	0.112	0.124	0.054
Adjusted R^2	0.492	0.108	0.120	0.050
Residual Std. Error (df = 1373997)	0.353	0.341	0.422	0.231

Note: *p<0.05; **p<0.01; ***p<0.001

Table EC19: Regression Results: Controlling for Cardiac-Related Comorbidities

	Dependent variable:			
	Dee sed	Bleed	Cardiac	Non GI
	(1)	(2)	(3)	(4)
Cardiac	0.002	0.001	0.005	0.007
	(0.009)	(0.004)	(0.011)	(0.005)
integ	−0.041***	0.001*	0.0004***	0.003*
	(0.008)	(0.0005)	(0.0001)	(0.002)
Observations	1,380,007	1,380,007	1,380,007	1,380,007
R^2	0.475	0.114	0.134	0.051
Adjusted R^2	0.472	0.110	0.130	0.047
Residual Std. Error (df = 1373997)	0.352	0.341	0.421	0.233
Note:				*p<0.05; **p<0.01; ***p<0.001