Brady Post (Orcid ID: 0000-0001-6544-8744)
Edward C. Norton (Orcid ID: 0000-0003-4555-0631)

Title page

Title: Hospital-Physician Integration and Risk-Coding Intensity

Running title: Vertical Integration and Risk Coding

Keywords: Physicians, Hospitals, Medicare, Vertical Integration

Hospital-Physician Integration and Risk-Coding Intensity

Abstract

Hospital-physician integration has surged in recent years. Integration may allow hospitals to share resources and management practices with their integrated physicians that increase the reported diagnostic severity of their patients. Greater diagnostic severity will increase practices' payment under risk-based arrangements. We offer the first analysis of whether hospital-physician integration affects providers' coding of patient severity. Using a two-way fixed effects model, an event study, and a stacked difference-in-differences analysis of 5 million patient-year observations from 2010-2015, we find that the integration of a patient's primary care doctor is associated with a robust 2-4% increase in coded severity, the risk-score equivalent of aging a physician's patients by 4-8 months. This effect was not driven by physicians treating different patients nor by physicians seeing patients more often. Our evidence is consistent with the hypothesis that hospitals share organizational resources with acquired physician practices to increase the measured clinical severity of patients. Increases in the intensity of coding will improve vertically-integrated practices' performance in alternative payment models and pay-for-performance programs while raising overall health care spending.

Keywords:

Physicians, Hospitals, Medicare, Vertical Integration, Healthcare spending, Professional Labor Markets

JEL Classification:

I110, I180, J440

1. Introduction

Hospital ownership of physician practices and hospital employment of physicians, often called hospital-physician integration or hospital-physician vertical integration, has increased rapidly in the last decade.(Baker, Bundorf, Devlin, & Kessler, 2018; Baker, Bundorf, & Kessler, 2014; Nikpay, Richards, & Penson, 2018; O'Malley, Bond, & Berenson, 2011; Physicians Advocacy Institute, 2019; Welch, Cuellar, Stearns, & Bindman, 2013) While integrated hospital-physician systems have enjoyed several advantages, including generous reimbursement from Medicare,(Chernew, 2021; Dranove & Ody, 2019; Post, 2021) foreclosed rivals,(Richards, Seward, & Whaley, 2020) and more negotiating power,(Neprash, Chernew, Hicks, Gibson, & McWilliams, 2015) researchers and policymakers have voiced concerns that hospital-physician integration has increased prices and spending without an offsetting increase in quality.(Lin, McCarthy, & Richards, 2021; Young, Zepeda, Flaherty, & Thai, 2021)

Simultaneously, a growing body of research has discovered that hospitals are quite agile in responding to incentives to maximize reimbursement by strategically coding patient symptoms into diagnosis codes reported on claims to insurers.(Bastani, Goh, & Bayati, 2019; Meyers, Mor, & Rahman, 2020) Such coding practices, while enabling hospitals to achieve better scores in risk-adjusted payment settings, have increased public spending on health care by billions of dollars.(Geruso & Layton, 2020) The two may be linked: by vertically integrating with physician practices, hospitals may be able to extend their coding resources to physicians. This implies that one potential effect of vertical integration is an increase in the coded severity of the patient population of integrated doctors, but this remains a gap in the current literature. Given the rapid rise of hospital-physician integration, the pervasive incentives to increase coded severity, and the growing policy concern attached to both, our focus in this paper is to evaluate whether increases in patients' coded severity of illness constitute one of the effects of hospital-physician integration.

1.1. Coded Severity and Vertical Integration

Hospital systems have strong financial interests in ensuring that patients within their systems appear to be as medically complex as possible. Hospitals may use integration with physicians – even physicians in outpatient settings – as a means to achieve this end. This could happen in at least two ways. First, it is well-known that payment rates to hospitals for Medicare Severity Diagnosis Related Groups (MS-DRGs) increase with coded severity. In 2017, the Centers for Medicare and Medicaid Services (CMS) reimbursed a coronary bypass without a major complication and comorbidity at an average rate of $23,406; if the hospital coded the procedure with a major complication and comorbidity, the rate rose to $34,825.(Medtronic, 2018) Hospitals can use integration to ensure that physicians in their systems, even in outpatient settings, thoroughly code each patient's comorbidities at each visit; this can help to specify more lucrative MS-DRGs when the patient is hospitalized at a later date. For example, with a patient's outpatient diagnostic history on file, a hospital can observe the patient's past outpatient diagnosis of Type 2 diabetes with hyperlipidemia and code MS-DRGs with a comorbidity rather than lower-revenue MS-DRGs without one. This incentivizes hospitals to ensure that physicians in the community document patient conditions thoroughly.

Secondly, hospitals increasingly bear risk in pay-for-performance care delivery models. These include CMS's Hospital Value Based Purchasing program, the Hospital Readmissions Reduction Program, Accountable Care Organizations (ACOs), and others. Most pay-for-performance and alternative payment models use a risk-adjusted calculation for reimbursement or for measurement of participants' performance. Within the Medicare Shared Savings (ACO) Program, for example, CMS uses Hierarchical Condition Categories (HCCs) to calculate the spending risk of each patient attributed to an ACO, and thereby to determine appropriate budget benchmarks. Across all such programs, hospitals' performance depends on physicians' coding of patient severity. This implies that upon joining a hospital system, an integrated physician may be encouraged or expected to code patient symptoms to the highest severity possible. Since HCCs and other risk scores are calculated using diagnoses found on both inpatient and outpatient claims, these incentives are important not just in inpatient settings, but in outpatient settings in which many integrated physicians practice. A hospital could improve its performance in an ACO simply by directing physicians to document greater patient severity at every office visit.

Operationally, hospitals are also much more likely to have access to comprehensive coding resources than the average physician practice. Hospitals often have dedicated billing departments, software, and consultants to comprehensively identify the highest allowable diagnosis code for a given set of patient symptoms. A cottage industry has emerged to support hospital systems in their efforts to maximize reimbursement. The American Academy of Professional Coders, for example, offers 28 certifications, including "professional service coding," "professional billing," and "clinical documentation." Hospitals often employ full-time medical coders, and virtually all hospitals use electronic health records (EHRs). A wide variety of proprietary medical coding software applications plug into EHRs to allow providers to automatically identify the highest appropriate reimbursement for their services. These resources are expensive for small businesses, and despite recent consolidation, most physician practices are still small businesses, often consisting of just a few physicians.(Muhlestein & Smith, 2016)

Given these operational factors, hospitals may believe that physicians are leaving money on the table by failing to maximize patient severity.

In addition to motive, there is opportunity: if the frequency of incorrect and fraudulent claims is any indicator, coded severity is relatively easy to manipulate. The Department of Health and Human Services's Office of Inspector General (OIG) recently found that nearly half of evaluation and management (E/M) encounters were incorrectly coded, and that "upcoded" (i.e. inappropriately high) E/M services cost Medicare $4.6 billion in 2010.(Levinson, Grant, Durley, Bessette, & Verges, 2014) Similarly, CMS found $28 billion in payments attributable to improper claims in 2019.(Centers for Medicare and Medicaid Services, 2019) An estimated 10,000 of 60,000 reimbursed claims for present-on-admission hospital infections were upcoded, costing Medicare $200 million.(Bastani et al., 2019) Recent research found that hospitals that offered Medicare Advantage plans coded their patients more aggressively.(Meyers et al., 2020) In view of these circumstances, policymakers have been skeptical that rapid increases in coded severity reflect actual changes in underlying patient condition. In Massachusetts, for example, the state's Health Policy Commission found evidence of widespread increases in coded severity from 2013 to 2017, the spending equivalent of adding "an additional 428,000 commercially insured Massachusetts residents with complex diabetes or 920,000 with cerebral palsy."(Massachusetts Health Policy Commission, 2019) The commissioners concluded that coding practices, not patient health, drove heightened coded severity.

1.2. Our Contribution

In view of the incentives to increase coded severity, the rationale for hospital-physician vertical integration to enable it, and the policy implications, the effects of integration on coded patient severity are a critical gap in the literature. In this study, we used national Medicare claims data over a six-year window to evaluate whether patients treated by vertically integrated physicians exhibited increases in their coded severity. We attributed patients to physicians and measured whether their physicians were vertically integrated with a hospital. We used a large panel that included nearly a million patients and over 75,000 physicians. We addressed issues of both patient selection and physician selection through a variety of methods, including the use of a balanced sample, physician-pair fixed effects and year fixed effects, an event study estimation strategy that allowed for treatment effects that varied over time, and verification of pre-period parallel trends. Following recent developments in the differential timing literature, we further supplemented our approach with a stacked difference-in-differences model to confirm the robustness of our findings. These techniques enabled us to examine the assumptions of our analysis and identify a credible reduced-form estimate of our effect. We identified a statistically significant 2-4 percent increase in coded patient severity associated with integration. We emphasize here that our paper does not aim to quantify changes in spending: risk scores are important on their own, and they do not translate directly to spending (they are composite measures of patient illness constructed from the diagnoses that providers report in insurance claims; they are used to adjust for spending risk in many programs, but there is not a one-to-one conversion). To provide context for the magnitude of integration's effects, we compare our estimated effect to the effects of aging: a 2-4 percent increase in coded patient severity is the risk-score equivalent of an integrated physician's patients suddenly becoming 4-8 months older. This empirical finding ties into pay-for-performance considerations: through hospital-physician integration, hospitals may improve their case mix measures and, therefore, their success in pay-for-performance programs, while also increasing Medicare spending through higher risk-adjusted payments.

As a secondary matter, we also sought to paint a more complete portrait by examining some of the mechanisms that may have generated an effect on coded patient severity. We consider three likely possibilities. First, integration could change the coded severity of a physician's patients by changing the "back office" coding system, e.g., using technology that standardizes and maximizes coding, which we have discussed. Second, vertically integrated physicians might prescribe more health care for their patients, whether as part of a quality-improvement initiative – for example, integration has been associated with greater use of care management protocols – or simply because they enjoy higher levels of outpatient reimbursement.(Bishop, Shortell, Ramsay, Copeland, & Casalino, 2016; Post, Norton, Hollenbeck, Buchmueller, & Ryan, 2021) Coding of patient chronic conditions could increase if patients see their physician more often. Third, sicker patients may seek out hospital-integrated physicians if they anticipate needing access to hospital-based services. If so, the average severity of an integrated physician's panel would increase even without changes in coding behavior. Our analytical approach sought to at least provide preliminary evidence about the likelihood of each of the first two mechanisms while we endeavored with our sample construction to rule out the third as a confounding factor.

This study adds to the vertical integration literature by articulating an additional advantage for hospitals and physicians that integrate. Our findings also provide policymakers with insights about the role of vertical integration in coded severity and the rising cost of care.

2. Existing Literature

A growing empirical literature has highlighted some of the effects of vertical integration.(Post, Buchmueller, & Ryan, 2018) Jung and colleagues found that integration was associated with more expensive treatment choices for patients with cancer.(Jung, Feldman, & Kalidindi, 2019) Richards and colleagues found evidence of anti-competitive vertical foreclosure in outpatient markets in Florida stemming from hospital-physician integration.(Richards et al., 2020) Baker and colleagues found that hospital ownership of physician practices steered patients toward higher-cost, lower-quality hospitals.(Baker, Bundorf, & Kessler, 2016) Capps, Dranove and Ody identified double-digit increases in commercial outpatient prices;(Capps, Dranove, & Ody, 2018); Lin, McCarthy, and Richards found increases in inpatient prices; and Koch, Wendling, and Wilson found large increases in spending among Medicare patients.(Koch, Wendling, & Wilson, 2017) Vertical integration may offer additional advantages as well, including a partial resolution of the principal-agent problem, in which integration more closely aligns physician incentives with hospital revenue-maximizing objectives; or the ability to collect additional revenue due to site-of-service payment differentials.(Chernew, 2021; Dranove & Ody, 2019; Post et al., 2021) Other work has found that integration was associated with modestly increased probability of cancer screenings and improvements in some quality scores.(Carlin, Dowd, & Feldman, 2015; Zepeda, Nyaga, & Young, 2020)

These studies illustrate that integration is a complex “treatment.” It involves changes in employer, organizational culture, internal resources, billing patterns, and more. The motivation to integrate likely differs from provider to provider: a small suburban hospital may have different goals than a large urban health system; physician organizations, too, have their own goals. Correspondingly, the effects of integration are also myriad. Each of the effects found in previous studies has specific patient welfare implications and policy responses. Our purpose in this study was not to assess the relative importance of these many factors, but rather, to simply examine the first-order question of whether there was evidence that untapped patient coding opportunities constituted an additional advantage of vertical integration.

3. Methods

3.1. Data, Sample, and Key Variables

To measure the effect of hospital-physician integration on patient coded severity, several important data elements are required: patient demographic information and diagnosis history over several years; physician integration status over several years; and a large sample to allow for sufficient variation. These data elements are present in the data sources that we used. We identified our sample using national Medicare claims from 2010-2015. We linked several data sources: the 20 percent sample of fee-for-service Medicare Provider Analysis and Review (MedPAR) and Carrier claim line items from 2010-2015, Medicare Data on Provider Practice and Specialty (MD-PPAS), the Medicare Master Beneficiary Summary file, the American Community Survey, and the Area Health Resources File. The Medicare claims and Master Beneficiary Summary files, available to us under a data use agreement, contained records of the health care services used by Medicare enrollees, along with diagnosis codes and demographic information of each patient; this rich data source enabled us to construct a detailed analytical file. MD-PPAS is a provider-level file that includes specialty and practice information for each provider (physicians, nurse practitioners, and others) that billed Medicare. The American Community Survey and Area Health Resources File are publicly available datasets that contain economic and health information calculated by county and other geographic units. This study was deemed exempt by our organization's institutional review board.

With the information available in the data sources above, we measured each patient's age, sex, race, and number of months dually eligible for Medicaid coverage. We attributed each patient to the primary care physician from whom he or she received the plurality of his or her care that year. We also measured area and market characteristics for each beneficiary. We measured the county-level rates of high school completion, four-year college completion, the market penetration of Medicare Advantage plans, and the percentage of the local population at or below 138 percent of the federal poverty level. In addition, we measured the hospital referral region Herfindahl-Hirschman Index for hospitals, based on patient volume from MedPAR claims, as well as for primary care physicians based on patient volume from Carrier claims. Finally, we measured physician age and sex. We report sample characteristics in Table 1.

We calculated whether a physician was vertically integrated with a hospital each year using a claims-based algorithm developed by Neprash and colleagues in their 2015 *JAMA Internal Medicine* article (Neprash et al., 2015), subsequently used by others in published work (Marchetti et al., 2021; Post et al., 2021) and validated elsewhere.(Ho, Tapaneeyakul, Metcalfe, Vu, & Short, 2020) For each physician, we measured the percentage of claims billed under Medicare's office place of service codes and hospital outpatient department (HOPD) place of service (Appendix 1). When at least 75 percent of these claims were billed under the HOPD code, we classified the physician as integrated (we varied this threshold and our results did not substantially change; the distribution is bimodal, Appendix 2). We calculated our exposure variable as whether a patient's attributed physician in each year was vertically integrated. In all specifications, we controlled for calendar year.

Our dependent variable was the Hierarchical Condition Category (HCC) risk score of each patient in each year. Condition category scores have been used as measures of coded severity in previous work.(Geruso & Layton, 2020; Li, Sukul, Nuliyalu, & Ryan, 2020; Markovitz et al., 2019) This score, which we calculated in accordance with CMS specifications, measures the spending risk of the patient, with a population mean set to 1.00.(Pope et al., 2004) Patients with higher risk scores have higher predicted spending. Its two components are demographic predictors (administratively determined information such as age, sex, and disability status), which cannot be manipulated by physicians or health systems, and provider-reported diagnosis codes, which can. Providers can report both the number and intensity of diagnosis codes for their patients (i.e., coded severity), which affects their reimbursement. HCC scores combine both the number and intensity of diagnosis codes into a single composite index, which makes them an excellent comprehensive measure of patient severity. An additional useful feature of HCC scores is their direct relevance in revenue and pay-for-performance: HCC scores are used explicitly in calculating ACO payments, Medicare Advantage reimbursement, and other contracts that put financial risk on providers. HCC scores are also the measure most appropriate to our primary care-focused sample: HCC scores are sensitive to the behavior of primary care physicians, because primary care physicians have tremendous influence in diagnoses documented in outpatient settings, and HCC scores include outpatient diagnoses. Primary care physicians, by contrast, do not directly influence MS-DRG coding in inpatient settings, so we do not use MS-DRGs as an outcome. While we expect that physicians operating in inpatient settings might, when classifying an MS-DRG, take notice of the diagnosis information collected by primary care physicians in outpatient settings, our study is not designed to test this directly. We acknowledge that there are implications of our work that extend beyond the scope of what we can do in this paper; examining inpatient coding practices is one of them. In the present paper, we focus on HCC scores since they have been used in other published work; reflect a composite index of both the number and intensity of a patient's full set of diagnosis codes each year, most of which occur in the outpatient setting in which primary care physicians operate; and directly affect pay-for-performance programs. For the identification of our effect, we required variation in HCC scores. We found that patients exhibited substantial variation in their HCC scores during the years of our study period (Figure 1).

Figure 1. Variation in coded severity within patients

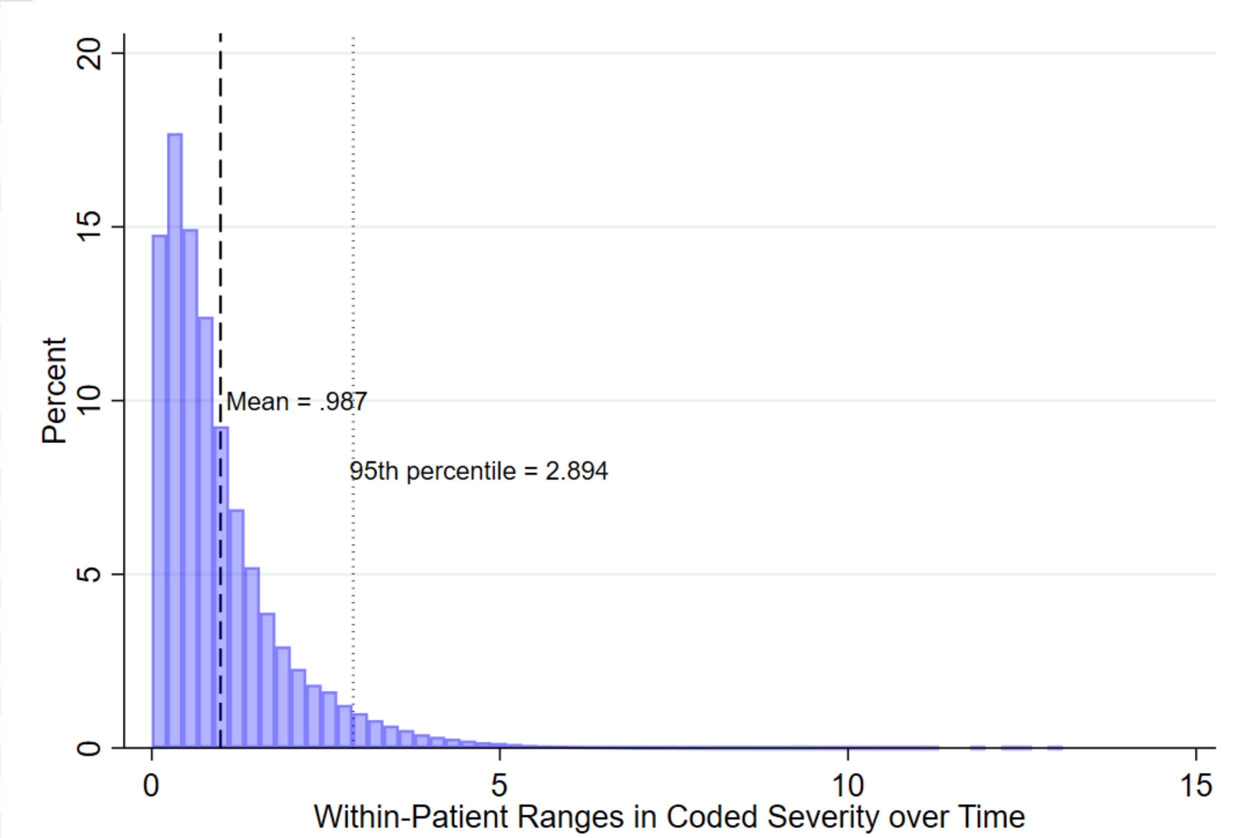


Figure 1 legend: HCC – Hierarchical Condition Category score. Coded severity defined as HCC score. Graph to illustrate variation in dependent variable. The average patient HCC score was 1.047. The average within-panel range in a patient's HCC score from 2010-2015 was 0.987; the 95th percentile of this within-panel range was 2.894.

3.2. Statistical Models

To identify our effect of interest, we used a two-way fixed effects model, an event study, and a stacked differences-in-differences model. This set of statistical approaches allowed us to assess the validity of our design and explore the assumptions of our modeling to better identify a robust effect. We further estimated several models to better understand potential mechanisms for our observed effect. We describe all of these models in further detail in this section.

The following quasi-experiment corresponds to our research question: some subset of the primary care physicians in our sample became integrated during the study period. The treatment group was this subset. We sought to test whether becoming integrated was associated with changes in primary care coding behavior. We excluded physicians who were integrated throughout 2010-2015 to establish temporal precedence. The control group was the subset of primary care physicians who were not integrated from 2010-2015.

This setting entails three main challenges to validity – patient selection, physician selection, and variation in treatment timing – each of which we sought to address. Since patients do not choose physicians randomly, selection could occur at the patient level; we addressed this through our sample construction and physician-patient pair fixed effects estimation (described next). Since physicians do not integrate randomly, selection could occur at the physician level; we describe how we handled this in the paragraphs following Equation (1). Since physicians became "treated" at different times, thereby raising concerns about measuring an average treatment effect in the presence of treatment effects that may be time-varying,(Goodman-Bacon, 2021) we examined effect sizes over time using an event study and a stacked difference-in-differences model, described in our discussion of Equation (2).

To manage the possibility of selection at the patient level, we constructed a balanced panel of primary care physician–patient pairs. Our large data set ensured sufficient sample to use a balanced panel. This approach captured changes in coded severity when a primary care physician became newly integrated among the set of patients he or she had already been seeing prior to integration. We used a PCP-patient (group) fixed effect for every pair: our identification came from changes in integration status and coded severity within a pair over time.

There are several advantages to this approach. Most importantly, it addresses the potential for endogeneity of integration and patient choice of doctor, i.e., selection at the patient level. A patient could, upon receiving some diagnosis, decide to switch into treatment by a vertically integrated primary care physician (perhaps anticipating need for future hospital-based care). Relatedly, upon integration, a PCP might begin seeing a new supply of patients as directed by the hospital, which could be, on average, more complex than those they had previously treated. These dynamics would imply real differences in the illness composition of a physician's panel (i.e., case mix) rather than changes in coding practices, the latter of which is our interest here. Our balanced pairs approach rules out the case mix problem and most closely aligns with the quasi-experiment above. The main disadvantage of this approach is that with many fixed effects, there is less variation available to identify an effect. We consider the benefits to be worth this cost; following Mummolo and Peterson's recommendations on transparency in fixed effects analysis, Appendix 3 shows the identifying variation that we used.(Mummolo & Peterson, 2018) We identified our effect from variation in n=24,690 physician-patient pairs (n=148,140 physician-patient-year observations over the full study period) with variation in integration status over time. Our identifying sample is comparable or larger in size to the identifying sample in other recent studies on hospital-physician integration.(Desai & McWilliams, 2018; Jung et al., 2019; Lin et al., 2021; Scott, Orav, Cutler, & Jha, 2017; Timbie et al., 2020)

We first estimated the following two-way fixed effects model:

$$(1) \qquad Severity_{ipt} = \delta_{ip} + \gamma Year_t + \alpha Integrated_{pt} + \beta X_{ipt} + \varepsilon_{ipt}$$

where $Severity_{ipt}$ denoted the coded severity (HCC score) for patient i attributed to physician p at year t. δ_{ip} was the group fixed effect for the patient-physician pair and $\gamma Year_t$ was the fixed effect for calendar year to control for secular trends (e.g. population age and clinical severity changes over time, independent of integration exposure). The variable of interest was $Integrated_{pt}$, which took the value of 1 if physician p was hospital-integrated in year t and 0 otherwise. X_{ipt} was a vector of time-varying patient and area characteristics including the area levels of Medicare Advantage penetration, completion of high school and four-year college, the percentage of the population at or below 138 percent of the federal poverty line, the market concentration of hospitals and primary care physicians; and patients' number of months dually eligible for Medicaid. Most of these covariates did not exhibit significant variation within physician-patient pairs, and thus a more parsimonious equation that excluded X_{ipt} was not materially different from our estimates; we report the estimates, which are slightly more conservative, from the simpler model (Table 2). ε_{ipt} was an error term clustered at the physician level. We report full regression results in Appendix 4.

Two-way fixed effects models are often viewed as difference-in-differences estimates, consisting of variance-weighted average treatment effects if the traditional parallel trends and exogeneity assumptions are met.(Goodman-Bacon, 2021) To gain insight into the pre-treatment trends of the treatment and control groups, we first estimated an event study model. We defined the control group as the set of physicians who did not integrate during our study window and the treatment group as the physicians who began unintegrated but became integrated during the study period. We defined each time period relative to the first year of integration (t = 0) for each physician. We estimated a model in which, rather than including a single treatment indicator (as with our two-way fixed effects model), we interacted treatment status with each relative time period to create a model fully saturated with treatment leads and lags, dropping the last pre-period.(Cunningham, 2021) We estimated:

$$(2)\ Severity_{ipt} = \delta_{ip} + \gamma Year_t + \tau_{pt} + \sum_{l=-5}^{4} \alpha_l\ 1(t - t_i^* = l)\ x\ Treatment_p + \beta X_{ipt} + \varepsilon_{ipt}$$

in which τ_{pt} controls for time relative to integration. $1(t - t_i^* = l)$ was an indicator for the time relative to the year of integration. $Treatment_p$ took the value of 1 if the physician ever integrated (i.e. $Treatment_p$ was stable within a physician panel). The treatment effects were contained in the vector α_l where $l \geq 0$. The other variables are constructed as in (1). ε_{ipt} was an error term clustered at the physician level. Figure 2 shows our results.

We took steps to address selection at the physician level. The validity of any difference-in-differences analysis could be undermined by uncontrolled factors that determine both treatment status and outcome of interest. Here, this could occur through selection at the physician level: physicians who become integrated might be different than physicians who do not due to factors correlated to coded patient severity (e.g., perhaps a physician's medical school influences their preference for integration as well as coding behavior). We addressed this potential concern in two ways. First, we followed the approach used in previous vertical integration literature by including physician and time fixed effects.(Koch et al., 2017) Doing so eliminates confounding from factors that are time-invariant (including a physician's sex, medical training, et al.), and leaves only factors that change over time and are correlated to both integration status and the outcome of interest. Second, mirroring recent work in this area,(Lin et al., 2021) we tested whether physicians who became integrated exhibited distinct trends in the pre-treatment period compared to the control physicians. We found no evidence for such pre-trends. Taken together, these suggest that the control group served as a valid counterfactual to the treatment group.

The last of the major threats to the validity of our analysis was variation in treatment timing. To this end, our event study specification was useful since it flexibly allows for differences in treatment effects over time. In addition, we supplemented our analyses with a stacked difference-in-differences model. Our setting was characterized by staggered treatment start dates: physicians entered into vertical integration at different times throughout our study period. The analytical implication is that there is no single starting date from which to measure time since integration.

A recent body of work has focused on potential issues related to differential timing in two-way fixed effects models and difference-in-differences designs.(Abraham & Sun, 2018; Athey & Imbens, 2018; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021) Cengiz, Dube, Lindner, and Zipperer (2019) estimated a stacked event-specific model, or a stacked difference-in-differences model, to confirm that their results were robust to the negative weighting pitfalls of staggered timing identified in Abraham and Sun (2018). Cengiz and colleagues sought to estimate the effect of the minimum wage, but legal changes to the minimum wage occurred at different times for different states, i.e., treatment was staggered.(Cengiz, Dube, Lindner, & Zipperer, 2019) Cengiz and colleagues created event-specific datasets comprised of the treated group and clean controls (states with no minimum wage changes during the event window) for each event with at least three pre-periods, then stacked these datasets to estimate an average effect. By aligning their datasets in event-time, their setting was equivalent to a treatment event that happened at one time; and dropping any control states with minimum wage changes during the event window reduced the bias associated with heterogeneous treatment (see Cengiz Appendix D for further details).

We followed the Cengiz method to address potential concerns about differential timing. We identified the set of physicians for whom we could identify three pre-periods and one post-period, which yielded three cohorts (physicians who became integrated in 2013, 2014, and 2015 and remained integrated thereafter). As in Cengiz 2019, we applied stringent criteria to the control group for each event, limiting it to those who were never integrated during each respective timeframe.(Cunningham, 2021) We calculated time relative to integration, dropping the last pre-treatment period; we appended these datasets and estimated a model with leads and lags using this sample, clustering standard errors at the physician level. These results further supported our analysis and are depicted in Appendix 5.

3.3. Mechanisms

Our primary objective in this study was to identify the effect of integration on the coding of patient severity. However, we also considered two potential mechanisms for how coding changes might occur. The first is a back-office change in coding practices. Based on anecdotes from conversations with physicians, integration with a hospital brings a measure of standardization to coding practices, e.g. by using a common set of software tools and other technology. This enables hospitals to code all patients comprehensively and systematically in their systems. If so, conditional on observables, the variance in coded severity should fall after integration. To test for evidence of such standardization, we estimated the following two-way fixed effects model, controlling for the mean of a physician's patient severity:

$$(3)\ \ StdDev_{pt} = \delta_p + \gamma Year_t + \alpha Integrated_{pt} + \ \beta AverageHCC_{pt} + \varepsilon_{pt}$$

in which $StdDev_{pt}$ is the standard deviation of the coded severity of a physician's patients in year *t*, δ_p is a physician fixed effect, $\gamma Year_t$ is a year fixed effect, and $Integrated_{pt}$ indicates whether a physician p was integrated in year *t*. $AverageHCC_{pt}$ controls for the average coded severity of a physician's panel in year *t* (to control for the relationship between mean and variance). We interpreted α as the effect of integration on the standardization of coding practices. We report these estimates in Table 3. We also estimated an event study version of this model analogous to that of Equation (2) (Figure 3).

Integration may also influence coded patient severity through utilization. Integrated delivery systems may seek to deliver on promises of coordinated care by improving patient engagement and encouraging patients to check in with their primary care physicians more frequently. If so, changes in risk score might simply reflect that physicians have spent more time with patients, i.e., have had more chances to notice comorbidities and code them. To account for this possibility, we sought to estimate a model that accounted for patient-physician visits. We calculated each patient's number of office visits in each year (with all physicians, not just with their main physician, since vertical integration may enable more in-house referrals). We added the number of office visits as a covariate to our two-way-fixed effects model and estimated:

$$(4)\ Severity_{ipt} = \delta_{ip} + \theta OfficeVisits_{pt} + \gamma Year_t + \alpha Integrated_{pt} + \beta X_{ipt} + \varepsilon_{ipt}$$

If integration increases coded severity though increased utilization, adding office visits to Equation (1) will reduce the estimated effect of integration, i.e., $\alpha_{Eq4} < \alpha_{Eq1}$. If increased utilization is not a mechanism, the estimate will remain unchanged. We report our results in Table 3.

4. Results

4.1. Descriptive

We identified a balanced analytical sample of 5,408,352 physician-patient-year observations, representing 901,392 unique patients who were attributed to 76,009 unique primary care physicians (Table 1). In 2010, 28,913 patients received care from integrated physicians; by 2015, this number had grown to 45,898.

The patients treated by integrated physicians closely resembled the patients treated by non-integrated physicians in demographics: in 2010, patients were approximately 72 years of age in both groups; about 57 percent of the sample was female; and the racial distribution was very similar. Consistent with our balanced sample, the average age at the end of the study had grown to approximately 77 years of age by 2015. Patients also appeared similar in area characteristics, such as the poverty rate, the percentage of the population that graduated from high school and college, and the Medicare Advantage penetration rate in the county. Hospital markets were slightly more concentrated among patients treated by integrated physicians (HHI of 2210 compared to 1930), a pattern also shown in the market for primary care physicians (178 compared to 150). The non-integrated physician population was slightly older (52 years old compared to 50) and had fewer females (19 percent compared to 30 percent) than the integrated physician population.

Patients were similar in their clinical profiles regardless of treatment from integrated or non-integrated physicians. Those treated by non-integrated physicians in 2010 were coded with an average of 1.56 chronic conditions, while patients treated by integrated physicians had an average of 1.53 chronic conditions. Similarly, the HCC coded severity score was nearly identical between integrated and non-integrated physicians in 2010 (0.90 in both groups). Consistent with an aging population, coded severity grew over time, with those treated by non-integrated physicians reaching an average of 1.26 and by integrated physicians reaching 1.25.

4.2. Statistical Results

4.2.1. Main models

One of the key assumptions underpinning this analysis was that the difference between the treatment group and the comparison group would have remained constant from the pre-period through the post-period in the absence of treatment (parallel trends). To justify this assumption, we examined our data for leading effects in the pre-treatment period. Figure 2 displays our event study. None of the leading effects of integration were statistically different from zero, which provided evidence in favor of our parallel trends assumption.

Figure 2. Event study of coded severity

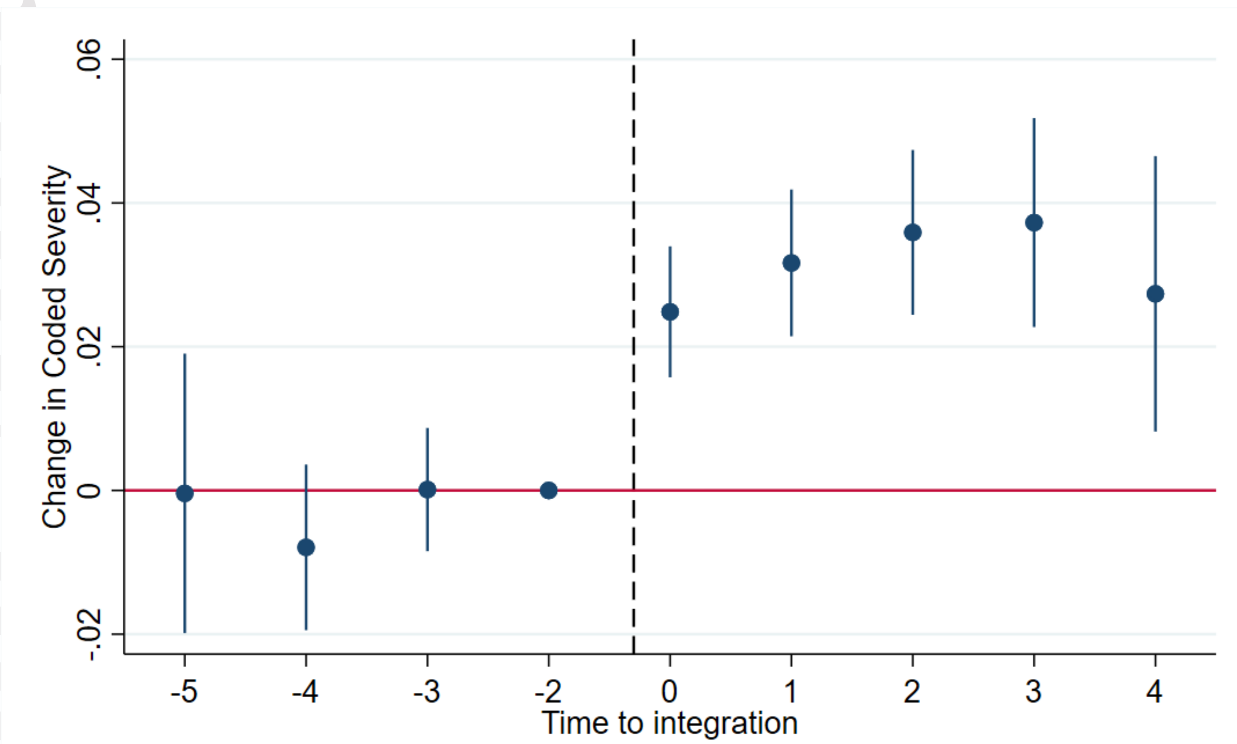


Figure 2 legend. This figure displays the event study specification of the effect of integration on the coded severity of an integrated physician's patients. The point estimates are the regression coefficients from Equation (2). The y-axis displays the dependent variable (natural log of Hierarchical Condition Category scores). The x-axis displays the time relative to when the physician became hospital-integrated (t = 0 represents the first year of integration). These estimates imply that patients' risk scores increased between 2-4 percent after their primary care physician vertically integrated with a hospital.

We proceeded to our two-way fixed effects model. The first column of Table 2 reports our preferred specification for the effect of integration on coded severity. Following prior literature, we expected key variables to have a proportional effect, and therefore measured our outcome variable in natural logs.(Lin et al., 2021) We identified a statistically significant and precisely estimated 2.02 percent increase in coded severity associated with integration. This result was robust to including a large set of covariates (second column).

In our event study specification, we identified statistically significant increases in patient coded severity in every post-period, ranging from about 2.5 percent in the first post-period to about 3.7 percent in the third post-period (Figure 2). We interpreted this as evidence that coded severity increased with integration.

To further probe the credibility of our design, we estimated our stacked difference-in-differences model. Each of the three events included three pre-periods and one post-period. The results from this analysis further bolstered our main estimates. As with our main event study, none of the leading effects were statistically distinguishable from zero. The first-period effect was equal to about 2.30 percent, closely matching the estimate from our first-period effect in our event study (Appendix 5).

Since HCC scores do not directly translate to spending changes, we provided practical context for the magnitude of integration's effects by comparing them to the effects of aging. In our sample, as patients aged, their HCC scores increased by about 5.8 percent per year. The effect sizes of integration that we identified ranged from about 35 percent of this value (2.02 / 5.8, from our two-way fixed effects model) to about 64 percent (3.7 / 5.8, from our event study). This implied that, roughly speaking, integration was associated with the risk-score equivalent of aging 4 to 8 months (0.35*12 to 0.64*12).

4.2.2. Mechanisms

Our conclusion that integration was associated with increases in coded severity is an important finding irrespective of the mechanism. However, we were also interested in whether this effect might have come from standardization of diagnosis coding protocols (e.g., hospitals sharing software and other resources) or by physicians having more opportunities to detect patient diseases (i.e., patients seeing physicians more often).

We first measured the standard deviation in patients' coded severity within physicians each year to gauge whether there was evidence of standardization of coding following integration. This analysis had less power to detect an effect than our main analysis, since our outcome was summarized at the physician-year level and therefore estimated with many fewer observations. Regardless, using a two-way fixed effects model, we found a statistically significant decrease in the standard deviation of a physician's coded severity scores associated with integration. Standard deviation changed by an average of -0.023; relative to a sample mean of 0.664, this was a decline of about 3.4 percent (first column of Table 3). We also estimated an event study (Figure 3). No leading effects were significant. Point estimates suggested an effect size of about -0.02, although some post-period effects were only marginally significant. Taken together, we considered these results suggestive, though not conclusive, of a standardization mechanism.

Figure 3. Event study of standard deviation in coded severity

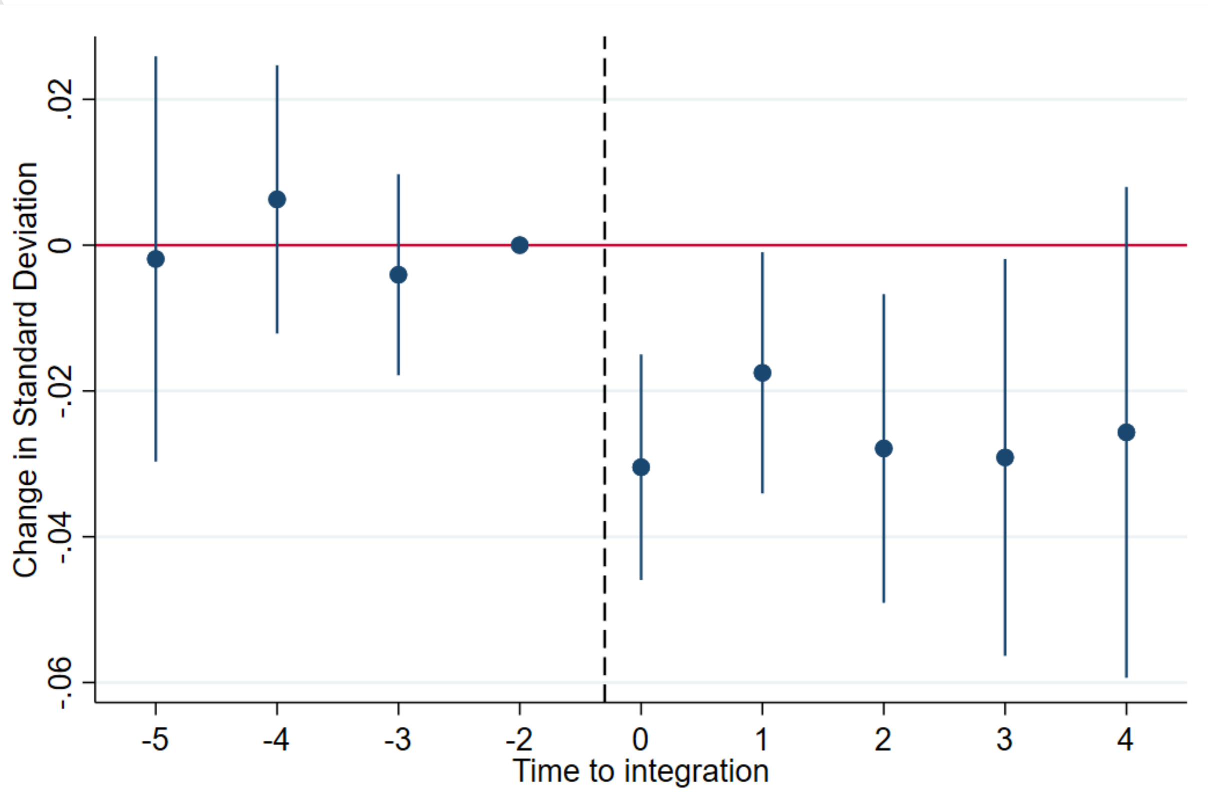


Figure 3 legend. This figure displays the event study specification of the effect of integration on the variability in risk scores within a physician's panel of patients. For each physician in each year, we calculated the standard deviation of his or her patient's Hierarchical Condition Category (HCC) scores. We then estimated a physician-year level event study model to determine whether integration exerted an influence on the variability of HCC scores. The sample mean of standard deviation in the last pre-period was 0.664. The point estimates are the regression coefficients from Equation (3). The y-axis displays the dependent variable (natural log of HCC scores). The x-axis displays the time relative to when the physician became hospital-integrated (t = 0 represents the first year of integration). These estimates demonstrate that the variability of HCC scores within a physician's panel of patients (even holding the panel of patients constant) decreased modestly after the physician vertically integrated with a hospital.

We then tested whether integration might increase patient coded severity via increasing patient-physician visits. We counted each patient's total office visits in each year and included it as a covariate in our two-way fixed effects model (second column of Table 3). As expected, the number of office visits had a positive and independent effect on severity, consistent with higherneeds patients seeing the doctor more often. However, including this control in our model did not, as the mechanism would predict, make the coefficient of interest smaller; our estimate of the effect of integration remained similar to that of our main specification. This implied that integration did not increase coded severity by increasing utilization of office visits.

5. Discussion

Using a nationwide six-year panel of over 900,000 patient-physician pairs, this study found that hospital-physician integration was associated with statistically and economically significant increases in patients' coded severity of illness. Our results were consistent with physician anecdotes that hospitals extend organizational resources such as clinical documentation support to acquired physician practices, which helps to increase the measured severity of acquired physicians' patients. The increases in severity were not driven by physicians seeing sicker patients nor by patients seeing physicians more often. Through higher coded severity, integration is likely to increase risk-adjusted health spending and improve provider performance in alternative payment and pay-for-performance models.

This study fills an important gap in the literature. While both hospital-physician integration and patient coded severity have been quickly rising – and along with them, health care spending – this study is the first to link the two. Our two-way fixed effects and event studies imply an effect size on the order of 2-4 percent. Though the relationship between coded severity and health care spending is not one to one, this effect magnitude is comparable to recent work on integration and prices (Lin et al., 2021) and spending.(Neprash et al., 2015)

The literature has not reached a consensus on the question of whether alternative payment models cause integration.(Alpert, Hsi, & Jacobson, 2017; Neprash, Chernew, & McWilliams, 2017; Ouayogodé, Fraze, Rich, & Colla, 2020) Our results do not answer this question, either. Instead, our results suggest that by extending coding-maximization protocols to physician practices, integration may help provider systems achieve reimbursement goals within alternative payment models.

With respect to patient welfare, our results can be seen in two ways. On one hand, these results look like opportunistic gaming of the payment system, in which the very same patients suddenly appear to be sicker than they really are. Related work and media coverage give credence: a policy change allowing more diagnosis codes resulted in higher reported patient illness; coding acrobatics accounted for a large percentage of apparent reductions in hospital readmissions; and a whistleblower in an integrated setting alleged outright coding fraud.(Ody, Msall, Dafny, Grabowski, & Cutler, 2019; Schulte, 2019; Sukul et al., 2019) We agree that this is a possible explanation, and if so, the implications are clear: hospital-physician integration boosts coded severity with no patient benefit and needlessly increases Medicare spending. A more charitable interpretation is that integration is helping to professionalize medicine. In this view, patients might finally be getting the diagnoses they should have gotten all along, and hospital resources that ensure more thorough symptom documentation improve patients' lives. One prediction from this latter interpretation is that the quality of care should improve with integration. Evidence for such improvement has remained limited.

Our data allow us to shed light on the relationship between integration and coded severity over nearly a million patients and over 70,000 primary care physicians over a six-year study window characterized by substantial integration activity. However, limitations remain. Our approach lacks true random assignment: while we have endeavored to mitigate selection bias, endogenous selection into integration remains a possible threat to validity. Further, while we observed that treatment effects were comparable regardless of year of integration, it is possible that hospitals selectively integrated practices that had lagged in coding sophistication; effects of integration in the future may differ. We study only the Medicare fee-for-service population and cannot infer the effects of integration on coded severity among the commercially-insured or other patient populations. We suspect, though, that if hospitals extend their coding resources to acquired practices, they likely do so for all patients treated by the practice. Additionally, coded patient severity matters most to providers operating in the context of Medicare Advantage, Accountable Care Organizations, and other risk-based models. Since the incentives to upcode in fee-for-service Medicare (our context) are not as strong, our results likely represent the lower bound of integration's effect on severity; extensions of our work could find even larger effects. Lastly, we lack data on underlying costs. Investigating this topic using detailed data on provider costs could persuasively answer questions about efficiencies and economies of scale. These are areas for future research.

Consolidation has been empirically tied to increases in prices and spending. Here, we tie it to increases in the reported illness of patients. As integrated delivery systems expand, and with them greater coded severity, state and federal policymakers tasked with constraining health care spending may need to rethink the role of provider-reported acuity in the reimbursement paradigm and in pay-for-performance models. We join other researchers in recommending more scrutiny of hospital-physician integration.

References

Abraham, S., & Sun, L. (2018). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3158747

Alpert, A., Hsi, H., & Jacobson, M. (2017). Evaluating the role of payment policy in driving vertical integration in the oncology market. *Health Affairs*, *36*(4), 680–688. https://doi.org/10.1377/hlthaff.2016.0830

Athey, S., & Imbens, G. W. (2018, August 15). Design-based analysis in difference-in-differences settings with staggered adoption. *ArXiv*. Retrieved from http://arxiv.org/abs/1808.05293

Baker, L. C., Bundorf, M. K., Devlin, A. M., & Kessler, D. P. (2018). Hospital Ownership of Physicians: Hospital Versus Physician Perspectives. *Medical Care Research and Review*, *75*(1), 88–99. https://doi.org/10.1177/1077558716676018

Baker, L. C., Bundorf, M. K., & Kessler, D. P. (2014). Vertical integration: hospital ownership of physician practices is associated with higher prices and spending. *Health Affairs (Project Hope)*, *33*(5), 756–763. https://doi.org/10.1377/hlthaff.2013.1279

Baker, L. C., Bundorf, M. K., & Kessler, D. P. (2016). The effect of hospital/physician integration on hospital choice. *Journal of Health Economics*, *50*, 1–8. https://doi.org/10.1016/j.jhealeco.2016.08.006

Bastani, H., Goh, J., & Bayati, M. (2019). Evidence of upcoding in pay-for-performance programs. *Management Science*, *65*(3), 1042–1060.

https://doi.org/10.1287/mnsc.2017.2996

Bishop, T. F., Shortell, S. M., Ramsay, P. P., Copeland, K. R., & Casalino, L. P. (2016). Trends in hospital ownership of physician practices and the effect on processes to improve quality. *The American Journal of Managed Care*, *22*(3), 172–176. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27023022

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. https://doi.org/10.1016/j.jhealeco.2018.04.001

Carlin, C. S., Dowd, B., & Feldman, R. (2015). Changes in quality of health care delivery after vertical integration. *Health Services Research*, *50*(4), 1043–1068. https://doi.org/10.1111/1475-6773.12274

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, *134*(3), 1405–1454. https://doi.org/10.1093/qje/qjz014

Centers for Medicare and Medicaid Services. (2019). *2019 Estimated Improper Payment Rates for Centers for Medicare & Medicaid Services (CMS) Programs*. Retrieved from https://www.cms.gov/newsroom/fact-sheets/2019-estimated-improper-payment-rates-centers-medicare-medicaid-services-cms-programs

Chernew, M. E. (2021). Disparities in payment across sites encourage consolidation. *Health Services Research*, *56*(1), 5–6. https://doi.org/10.1111/1475-6773.13612

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

Desai, S., & McWilliams, J. M. (2018). Consequences of the 340B drug pricing program. *New England Journal of Medicine*, *378*(6), 539–548. https://doi.org/10.1056/NEJMsa1706475

Dranove, D., & Ody, C. (2019). Employed for Higher Pay? How Medicare Payment Rules Affect Hospital Employment of Physicians. *American Economic Journal: Economic Policy*, *11*(4), 249–271. Retrieved from https://www.aeaweb.org/articles?id=10.1257/pol.20170020

Geruso, M., & Layton, T. (2020). Upcoding: Evidence from medicare on squishy risk adjustment. *Journal of Political Economy*, *128*(3), 984–1026. https://doi.org/10.1086/704756

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2021.03.014

Ho, V., Tapaneeyakul, S., Metcalfe, L., Vu, L., & Short, M. (2020). Using Medicare data to measure vertical integration of hospitals and physicians. *Health Services and Outcomes Research Methodology*, *20*(1), 1–12. https://doi.org/10.1007/s10742-020-00207-7

Jung, J., Feldman, R., & Kalidindi, Y. (2019). The impact of integration on outpatient chemotherapy use and spending in Medicare. *Health Economics*, *28*(4), 517–528. https://doi.org/10.1002/hec.3860

Koch, T. G., Wendling, B. W., & Wilson, N. E. (2017). How Vertical Integration Affects the Quantity and Cost of Care for Medicare Beneficiaries. *Journal of Health Economics*, *52*, 19–32. https://doi.org/10.1016/j.jhealeco.2016.12.007

Levinson, D. R., Grant, D., Durley, J., Bessette, R., & Verges, M. (2014). *Improper Payments for Evaluation and Management Services Cost Medicare Billions In 2010*. Retrieved from https://oig.hhs.gov/oei/reports/oei-04-10-00181.asp

Li, J., Sukul, D., Nuliyalu, U., & Ryan, A. M. (2020). Patient Coded Severity and Payment Penalties Under the Hospital Readmissions Reduction Program. *Medical Care*, *58*(11), 1022–1029. https://doi.org/10.1097/MLR.0000000000001396

Lin, H., McCarthy, I. M., & Richards, M. (2021). Hospital Pricing Following Integration with Physician Practices. *Journal of Health Economics*, *77*. https://doi.org/10.1016/j.jhealeco.2021.102444

Marchetti, K. A., Oerline, M., Hollenbeck, B. K., Kaufman, S. R., Skolarus, T. A., Shahinian, V. B., … Modi, P. K. (2021). Urology Workforce Changes and Implications for Prostate Cancer Care Among Medicare Enrollees. *Urology*. https://doi.org/10.1016/j.urology.2020.12.051

Markovitz, A. A., Hollingsworth, J. M., Ayanian, J. Z., Norton, E. C., Moloci, N. M., Yan, P. L., & Ryan, A. M. (2019). Risk adjustment in Medicare ACO program deters coding increases but may lead ACOs to drop high-risk beneficiaries. *Health Affairs*, *38*(2), 253–261. https://doi.org/10.1377/hlthaff.2018.05407

Massachusetts Health Policy Commission. (2019). *Health Policy Commission Board Meeting*. Retrieved from https://www.mass.gov/files/documents/2019/09/11/20190911_Presentation_vFinal to post_0.pdf

Medtronic. (2018). *2018 Selected Cardiothoracic Procedures Coding Resource*. Minneapolis, MN.

Meyers, D. J., Mor, V., & Rahman, M. (2020). Provider Integrated Medicare Advantage Plans Are Associated With Differences In Patterns Of Inpatient Care. *Health Affairs*, *39*(5), 843–851. https://doi.org/10.1377/hlthaff.2019.00678

Muhlestein, D. B., & Smith, N. J. (2016). Physician Consolidation: Rapid Movement From Small To Large Group Practices, 2013–15. *Health Affairs*, *35*(9), 1638–1642. https://doi.org/10.1377/hlthaff.2016.0130

Mummolo, J., & Peterson, E. (2018). Improving the Interpretation of Fixed Effects Regression Results. *Political Science Research and Methods*, *6*(4), 829–835. https://doi.org/10.1017/psrm.2017.44

Neprash, H. T., Chernew, M. E., Hicks, A. L., Gibson, T., & McWilliams, J. M. (2015). Association of Financial Integration Between Physicians and Hospitals With Commercial Health Care Prices. *JAMA Internal Medicine*, *175*(12), 1–8. https://doi.org/10.1001/jamainternmed.2015.4610

Neprash, H. T., Chernew, M. E., & McWilliams, J. M. (2017). Little evidence exists to support the expectation that providers would consolidate to enter new payment models. *Health Affairs*, *36*(2), 346–354. https://doi.org/10.1377/hlthaff.2016.0840

Nikpay, S. S., Richards, M. R., & Penson, D. (2018). Hospital-Physician Consolidation Accelerated In The Past Decade In Cardiology, Oncology. *Health Affairs*, *37*(7), 1123–1127. https://doi.org/10.1377/hlthaff.2017.1520

O'Malley, A. S., Bond, A. M., & Berenson, R. A. (2011). Rising hospital employment of physicians: better quality, higher costs? *Issue Brief (Center for Studying Health System Change)*, (136), 1–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21853632

Ody, C., Msall, L., Dafny, L. S., Grabowski, D. C., & Cutler, D. M. (2019). Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated. *Health Affairs*, *38*(1), 36–43. https://doi.org/10.1377/hlthaff.2018.05178

Ouayogodé, M. H., Fraze, T., Rich, E. C., & Colla, C. H. (2020). Association of Organizational Factors and Physician Practices' Participation in Alternative Payment Models. *JAMA Network Open*, *3*(4). https://doi.org/10.1001/jamanetworkopen.2020.2019

Physicians Advocacy Institute. (2019). *Updated Physician Practice Acquisition Study: National and Regional Employment Changes in Physician Employment 2012-2018*. Retrieved from http://www.physiciansadvocacyinstitute.org/Portals/0/assets/docs/021919-Avalere-PAI-Physician-Employment-Trends-Study-2018-Update.pdf?ver=2019-02-19-162735-117

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., … Robst, J. (2004, June). Risk adjustment of medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, Vol. 25, pp. 119–141. Retrieved from /pmc/articles/PMC4194896/

Post, B. (2021). Association of a Medicare Outpatient Payment Reform with Hospital-Primary Care Integration: Heterogeneity Across Markets and Physicians (in press). *Medical Care, Published*(9). Retrieved from https://pubmed.ncbi.nlm.nih.gov/34593710/

Post, B., Buchmueller, T., & Ryan, A. M. (2018). Vertical Integration of Hospitals and Physicians: Economic Theory and Empirical Evidence on Spending and Quality. *Medical*

Care Research and Review, Vol. 75, pp. 399–433.
https://doi.org/10.1177/1077558717727834

Post, B., Norton, E. C., Hollenbeck, B., Buchmueller, T., & Ryan, A. M. (2021). Hospital-physician integration and Medicare's site- based outpatient payments. *Health Services Research*, *56*(1), 7–15. https://doi.org/10.1111/1475-6773.13613

Richards, M., Seward, J., & Whaley, C. (2020). *Treatment consolidation after vertical integration: Evidence from outpatient procedure markets*.
https://doi.org/10.7249/WRA621-1

Schulte, F. (2019, October 18). Whistleblower Accuses Seattle's Group Health Medicare Advantage Plan Of Fraud. *Shots - Health News: National Public Radio*. Retrieved from https://www.npr.org/sections/health-shots/2019/10/18/770466908/whistleblower-alleges-fraud-at-a-large-medicare-advantage-plan-in-seattle

Scott, K. W., Orav, E. J., Cutler, D. M., & Jha, A. K. (2017). Changes in Hospital–Physician Affiliations in U.S. Hospitals and Their Effect on Quality of Care. *Annals of Internal Medicine*, *166*(1), 1. https://doi.org/10.7326/M16-0125

Sukul, D., Hoffman, G. J., Nuliyalu, U., Adler-Milstein, J. R., Zhang, B., Dimick, J. B., & Ryan, A. M. (2019). Association Between Medicare Policy Reforms and Changes in Hospitalized Medicare Beneficiaries' Severity of Illness. *JAMA Network Open*, *2*(5), e193290.
https://doi.org/10.1001/jamanetworkopen.2019.3290

Timbie, J. W., Kranz, A. M., DeYoreo, M., Eshete-Roesler, B., Elliott, M. N., Escarce, J. J., … Damberg, C. L. (2020). Racial and ethnic disparities in care for health system-affiliated physician organizations and non-affiliated physician organizations. *Health Services*

Research, *55*(S3), 1107–1117. https://doi.org/10.1111/1475-6773.13581

Welch, W. P., Cuellar, A. E., Stearns, S. C., & Bindman, A. B. (2013). Proportion of physicians in large group practices continued to grow in 2009-11. *Health Affairs (Project Hope)*, *32*(9), 1659–1666. https://doi.org/10.1377/hlthaff.2012.1256

Young, G. J., Zepeda, E. D., Flaherty, S., & Thai, N. (2021). Hospital Employment Of Physicians In Massachusetts Is Associated With Inappropriate Diagnostic Imaging. *Health Affairs*, *40*(5), 710–718. https://doi.org/10.1377/hlthaff.2020.01183

Zepeda, E. D., Nyaga, G. N., & Young, G. J. (2020). The Effect of Hospital-Physician Integration on Operational Performance: Evaluating Physician Employment for Cardiovascular Services. *Decision Sciences*, *51*(2), 282–316. https://doi.org/10.1111/deci.12401

Tables

Table 1. Sample characteristics

	2010		2015	
	Not integrated	Integrated	Not integrated	Integrated
Unique Physician-Patient Pairs	872,479	28,913	855,494	45,898
Unique physicians	71,742	4,267	70,081	5,928
Unique patients	872,479	28,913	855,494	45,898
Patient Characteristics				
Age	72.2	71.6	77.2	76.8
Female (%)	57	57	57	57
White (%)	88	85	88	87
Black (%)	7	9	7	8
Hispanic (%)	1	1	1	1
Other race/ethnicity (%)	4	5	4	4
Number of chronic conditions	1.56	1.53	2.23	2.2
Average coded severity score	0.90	0.90	1.26	1.25
Area Characteristics				
Income below 138% of poverty level (%), county	13.7	14.2	10.6	10.6
High School (%), county	86.4	87.5	87.7	89
College (%), county	28	28.3	29.4	29.2
Medicare Advantage penetration (%), county	21.1	21.7	27.7	28.2
HHI Hospital, hospital referral region	1930	2210	2054	2333
HHI Primary Care, hospital referral region	150	178	155	176

Physician Characteristics				
Age	52	50.2	57	55.6
Female (%)	19	30	19	26

Table 1 notes: HHI – Herfindahl-Hirschman Index. Sample size across all years = 5,408,352 physician-patient-year observations, comprised of 901,392 unique patients and 76,009 unique physicians.

Table 2. Two-way fixed effects estimates: coded severity

	Coded severity (1)	Coded severity (covariates) (2)
Integration	0.0202*** (0.00288)	0.0206*** (0.00290)
2010 (omitted)		
2011	0.0550*** (0.000569)	0.0542*** (0.000571)
2012	0.107*** (0.000637)	0.106*** (0.000669)
2013	0.165*** (0.000676)	0.165*** (0.000771)
2014	0.228*** (0.000725)	0.228*** (0.000929)
2015	0.326*** (0.000788)	0.324*** (0.00108)
Covariates	No	Yes
Physician-patient pair fixed effects	Yes	Yes
Constant	-0.323*** (0.000468)	-0.308*** (0.0199)
Observations	5,408,352	5,408,352
R-squared	0.741	0.741

Table 2 notes: Standard errors reported in parentheses (clustered at physician level). *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports our main two-way fixed effects estimate from Equation 1. Column 2 adds covariates to the model; the estimated coefficient of interest remains approximately the same. The dependent variable is the natural log of a patient's coded severity (HCC score). Integration is a binary variable taking the value of 1 when a physician is integrated and 0 otherwise.

Table 3. Two-way fixed effects estimates: mechanisms

	(1) Standard deviation of coded severity	(2) Coded severity, controlling for number of office visits
Integration	-0.0227*** (0.00713)	0.0240*** (0.00272)
Number of office visits	N/A	0.0263*** (9.55e-05)
Covariates	No	No
Pair fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Control for average severity	Yes	No
Observations	374,717	5,408,352
R-squared	0.813	0.768

Table 3 notes: Standard errors reported in parentheses (clustered at physician level). *** p<0.01, ** p<0.05, * p<0.1. Physician-patient pair fixed effects and calendar year fixed effects included in all regressions. Column 1 reports the estimated effect of integration on the standard deviation of a physician's coded severity and includes a control for the average level of severity; results demonstrate that a physician's variance in risk scores decreased modestly (about 3.4 percent) after integration. Column 2 reports the results of Equation (4): it reports the estimated effect of integration on coded severity (natural log of HCC score) with an added control for the number of patient's office visits (a modification of Equation 1). Upon including a control for the number of office visits, the coefficient of interest is similar (even slightly larger) to the main two-way fixed effects specification, demonstrating that increased visits with physicians was not the mechanism for the increase in severity (which would imply a decrease in the coefficient of interest). Integration is a binary variable taking the value of 1 when a physician is integrated and 0 otherwise.

Figure Legends

Figure 1 legend: HCC – Hierarchical Condition Category score. Coded severity defined as HCC score. Graph to illustrate variation in dependent variable. The average patient HCC score was 1.047. The average within-panel range in a patient's HCC score from 2010-2015 was 0.987; the 95th percentile of this within-panel range was 2.894.

Figure 2 legend. This figure displays the event study specification of the effect of integration on the coded severity of an integrated physician's patients. The point estimates are the regression coefficients from Equation (2). The y-axis displays the dependent variable (natural log of Hierarchical Condition Category scores). The x-axis displays the time relative to when the physician became hospital-integrated (t = 0 represents the first year of integration). These estimates imply that patients' risk scores increased between 2-4 percent after their primary care physician vertically integrated with a hospital.

Figure 3 legend. This figure displays the event study specification of the effect of integration on the variability in risk scores within a physician's panel of patients. For each physician in each year, we calculated the standard deviation of his or her patient's Hierarchical Condition Category (HCC) scores. We then estimated a physician-year level event study model to determine whether integration exerted an influence on the variability of HCC scores. The sample mean of standard deviation in the last pre-period was 0.664. The point estimates are the regression coefficients from Equation (3). The y-axis displays the dependent variable (natural log of HCC scores). The x-axis displays the time relative to when the physician became hospital-integrated (t = 0 represents the first year of integration). These estimates demonstrate that the variability of HCC scores within a physician's panel of patients (even holding the panel of patients constant) decreased modestly after the physician vertically integrated with a hospital.

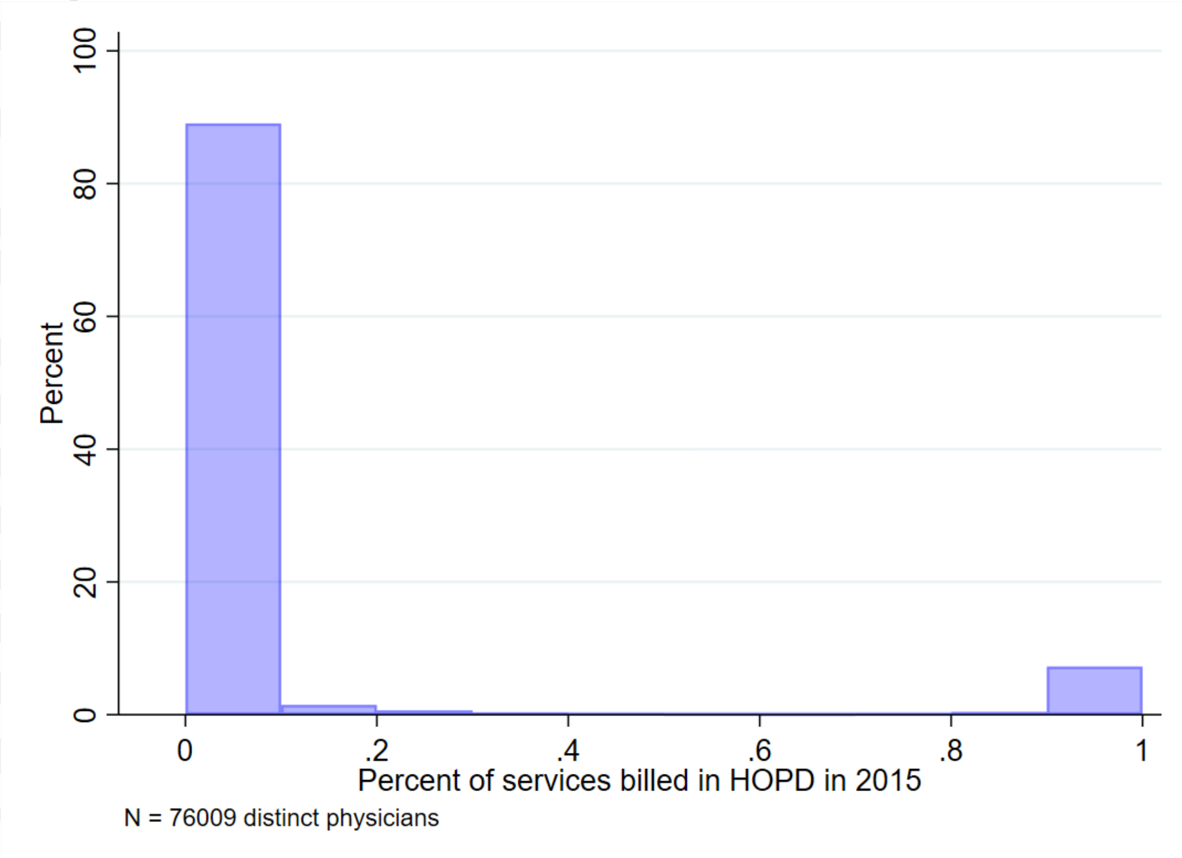
Appendices

Appendix 1. Independent variable (integration) and sample construction

We defined integration using a strategy developed by Neprash and colleagues (2015). They utilize the place of service codes found in Medicare claims. During our study period, when a hospital acquired a physician, the physician became eligible to bill under the Hospital Outpatient Department (HOPD) place of service code. The incentives to make this billing change were strong since reimbursement under an HOPD designation was higher than under an office designation. We created an indicator for each physician in each year to indicate integration status by counting the number of line items billed in the Medicare Carrier (physician/supplier) claims under an office code and under an HOPD code. When 75 percent or more of these line items were billed under an HOPD code, we classified the physician as integrated; Appendix 2 shows the distribution.

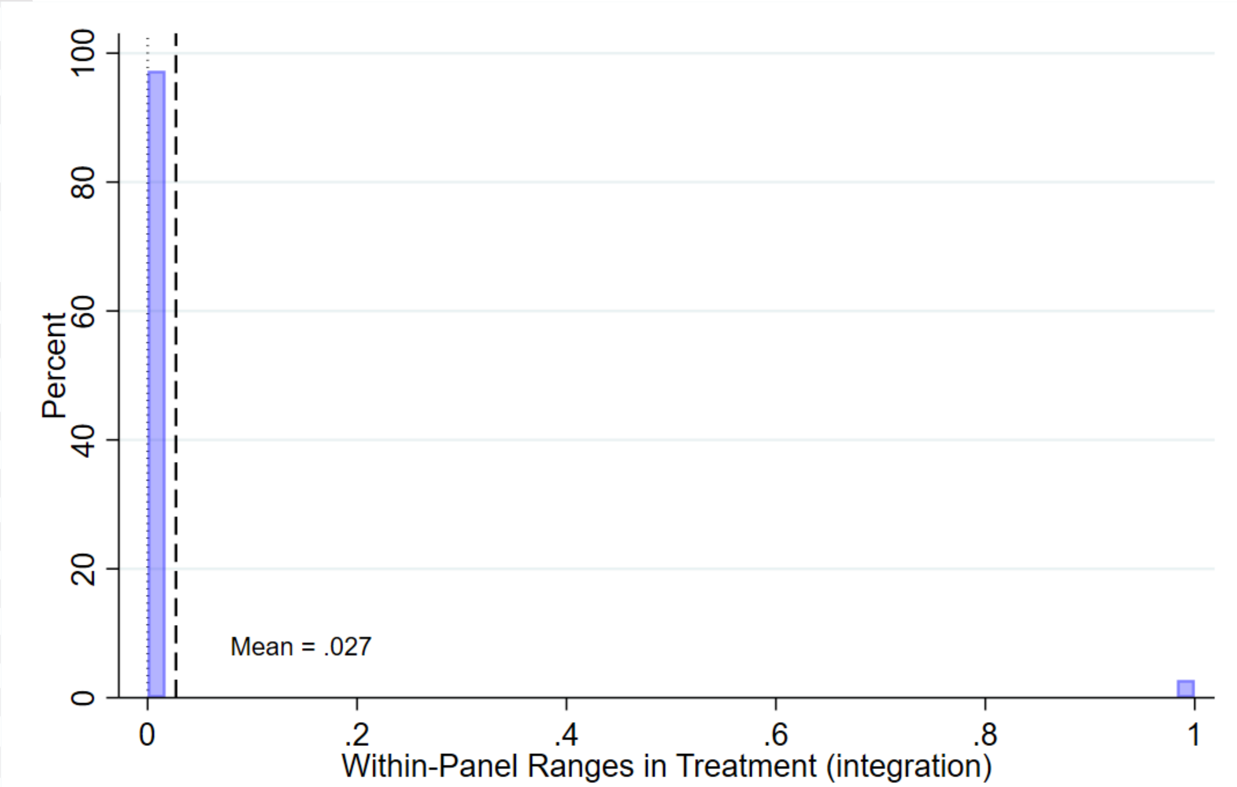
Eligible patients were those for whom data was sufficiently complete to calculate a Hierarchical Condition Category score during the study period and who were paired with a primary care physician in each year of the study period (n=5,828,802 physician-patient-years). We excluded a small number of patients for whom covariates were not present. Our final sample size was 5,408,352 physician-patient-years, corresponding to 76,009 unique physicians and 901,392 unique patients.

Appendix 2. Percent of Outpatient Services Billed from Hospital Outpatient Department Place of Service Codes in 2015



Notes: HOPD – hospital outpatient department. This graph plots one observation per physician (n = 76,009). Physicians with 75 percent or more of their outpatient services billed from an HOPD were defined as hospital-integrated (Neprash 2015). The distribution for other sample years is similarly bimodal.

Appendix 3. Within-Panel Ranges in Treatment (Identifying Variation)



Note: This graph plots one observation per physician-patient pair (n = 901,392). Identification comes from those who changed integration status (i.e., a range of 1, n = 24,688 pairs, or 148,128 pair-year observations.).

Appendix 4a. Two-way fixed effects regression

VARIABLES	(1) Preferred	(2) With Covariates
Integrated	0.0202***	0.0206***
	(0.00288)	(0.00290)
2011.year	0.0550***	0.0542***
	(0.000569)	(0.000571)
2012.year	0.107***	0.106***
	(0.000637)	(0.000669)
2013.year	0.165***	0.165***
	(0.000676)	(0.000771)
2014.year	0.228***	0.228***
	(0.000725)	(0.000929)
2015.year	0.326***	0.324***
	(0.000788)	(0.00108)
Number of dually-eligible months		0.0118***
		(0.000254)
Medicare Advantage % county penetration		5.73e-05
		(0.000103)
Percent county with high school		8.95e-06
		(0.000143)
Percent county with college		-2.34e-05
		(9.30e-05)
Percent county below 138 percent of poverty level		-0.000293***
		(0.000103)
Ln(Hospital HHI)		-0.00531**
		(0.00222)
Ln(PCP HHI)		0.00176**
		(0.000849)
Constant	-0.323***	-0.308***
	(0.000468)	(0.0199)
Physician-patient pair fixed effects	Yes	Yes
Clustered SE	Physician	Physician
Observations	5,408,352	5,408,352
R-squared	0.741	0.741

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports a parsimonious two-way fixed effects model. Column 2 adds covariates. Standard errors clustered at the physician level.

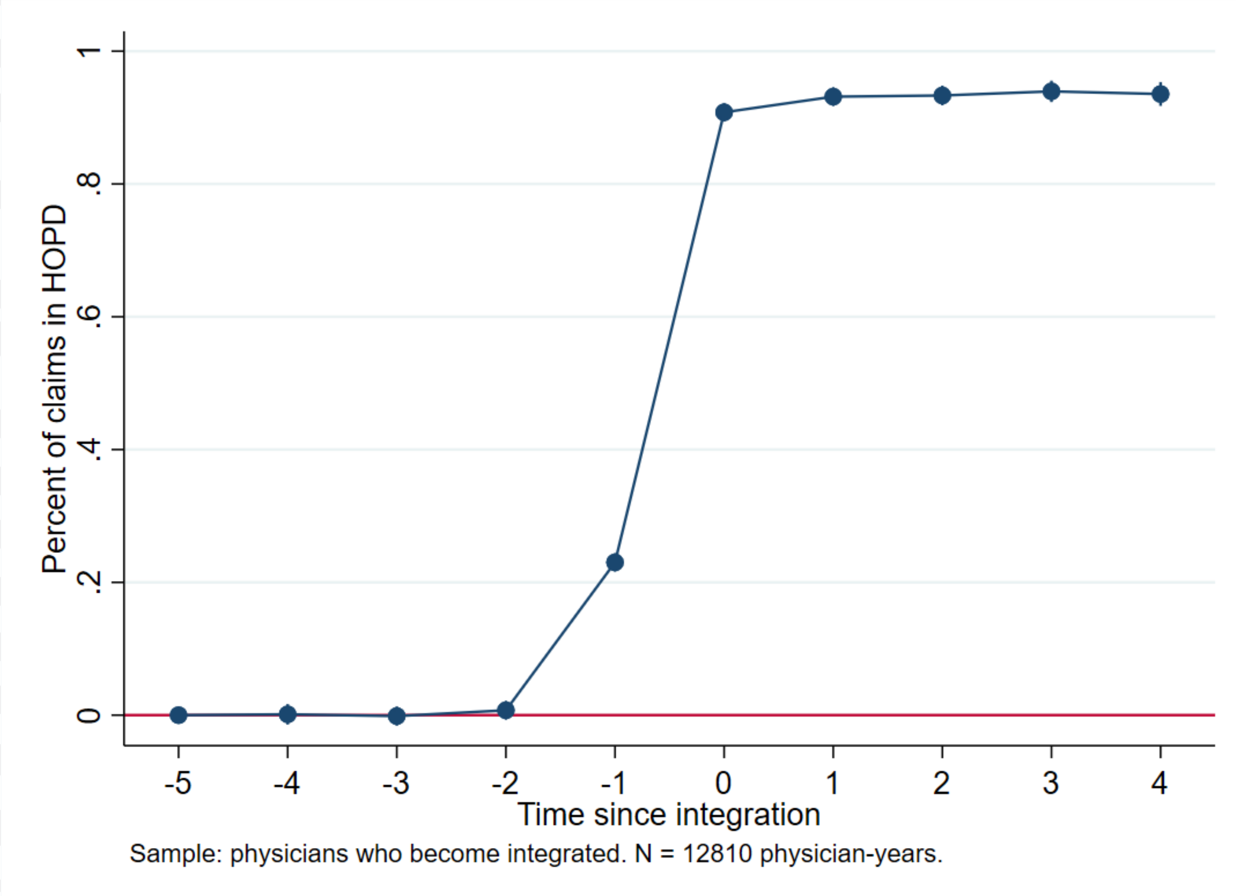
Appendix 4b. Event study regression

VARIABLES	Ln(HCC)	StdDev
t-5	-0.000405	0.00402
	(0.00992)	(0.0136)
t-4	-0.00792	0.00909
	(0.00588)	(0.00934)
t-3	0.000120	-0.000598
	(0.00437)	(0.00697)
t-2	Omitted	Omitted
t0	0.0249***	-0.0265***
	(0.00465)	(0.00795)
t1	0.0317***	-0.0126
	(0.00521)	(0.00845)
t2	0.0359***	-0.0224**
	(0.00585)	(0.0110)
t3	0.0373***	-0.0257*
	(0.00742)	(0.0141)
t4	0.0274***	-0.0211
	(0.00978)	(0.0178)
2011.year	0.0551***	-0.0240***

	(0.000584)	(0.00103)
2012.year	0.107***	-0.0447***
	(0.000655)	(0.00117)
2013.year	0.165***	-0.0690***
	(0.000695)	(0.00134)
2014.year	0.228***	-0.0941***
	(0.000747)	(0.00159)
2015.year	0.326***	-0.117***
	(0.000812)	(0.00210)
Constant	-0.323***	-0.363***
	(0.000477)	(0.00527)
Average HCC	N/A	1.036***
		(0.00563)
Physician-patient pair fixed effects	Yes	No
Physician fixed effects	No	Yes
Clustered SE	Physician	Physician
Observations	5,208,673	374,717
R-squared	0.740	0.813

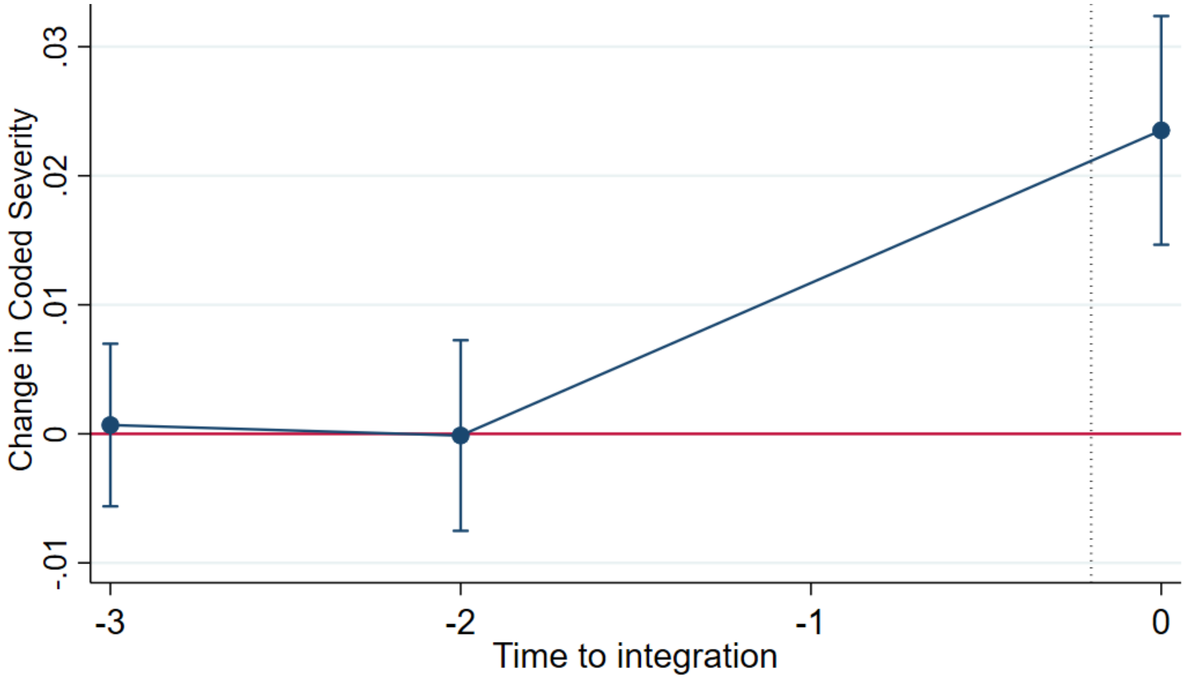
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports the results from our event study (equation 2). Column 2 reports an event study collapsed at the physician level with the standard deviation of a physician's HCC scores (i.e. across their patients in a year) as the dependent variable. Standard errors clustered at the physician level. Leading effects appear in parameter estimates for t-5 through t-2. Time t-1 was a contaminated year (both treated and control) because the definition of integration relied on calendar year activity (and some physicians who had integrated in, e.g., September did not qualify as "integrated" for the year). Appendix 4c shows this dynamic; for this reason, time t-1 is dropped from the event study specifications.

Appendix 4c. Time period preceding integration event as partially-integrated



Note: HOPD – hospital outpatient department. This graph depicts, among the integrating physicians, their share of services performed in the HOPD, which is used to define integration status each year. This graph shows that some physicians, though identified as not integrated (i.e. they are in t = -1) for the year, have spent at least part of the year integrated, thereby making them partially integrated. The event study analyses in this paper exclude t = -1 for this reason. However, including this time period and re-specifying Equation 2 such that it is considered the first post-period does not substantially change our findings; the first period has, as one would expect, a smaller treatment effect, but our conclusions are not materially affected.

Appendix 5. Stacked event study of coded severity



Appendix 5 legend. This figure displays the regression coefficients from a stacked event-specific model (also called a stacked difference-in-differences model) of the effect of integration on coded severity (Cengiz, Dube, Lindner, & Zipperer, 2019). The sample includes physicians for whom we could identify three pre-periods and one post-period. The y-axis displays the dependent variable (natural log of Hierarchical Condition Category scores). The x-axis displays the time relative to when the physician became hospital-integrated (t = 0 represents the first year of integration). These results imply robustness of our findings; the first-period estimate (2.3 percent increase) identified here is very similar to the first-period estimate from our main event study model (2.5 percent increase).