THE EFFECTS OF PHYSICIAN AND HOSPITAL INTEGRATION ON MEDICARE BENEFICIARIES' HEALTH OUTCOMES

Thomas G. Koch, Brett W. Wendling, and Nathan E. Wilson*

Abstract—We consider whether hospital acquisitions of physicians lead to improved clinical outcomes for Medicare patients aged 65 and older. The analysis combines 2005–2012 Medicare fee-for-service and enrollment data with merger and physician affiliation information from the Levin Reports and SK&A, respectively. We determine the effect of acquisitions on several health outcomes: mortality, acute myocardial infarctions, acute circulatory conditions, ischemic heart disease, glaucoma, symptomatic diabetes complications, and asymptomatic diabetes complications. These outcomes represent the progression of hypertension and diabetes into worse health states. Our results indicate that hospital acquisitions of existing physician practices have little effect on the health outcomes we consider.

I. Introduction

HEALTH care provider markets have experienced profound organizational changes since the turn of the century. For example, recent integration of physicians into hospital systems has led the proportion of U.S. physicians employed by hospitals to nearly double from 16% to 29% between 2007 and 2013 (Kane & Emmons, 2013).[1] This type of vertical integration may theoretically result in more efficient care and service. Indeed, Vita (2000) shows that vertical integration sometimes eliminates double marginalization, and Jin and Leslie (2009) find that it fosters quality. However, vertical integration could also lower incentives to innovate and result in higher prices.[2]

Policies addressing vertical integration in health care, and across other industries, depend on an understanding of their effects. The ambiguity in effects has led to growth in empirical research analyzing the effects of integration, and particularly health care provider integration.[3] Key research challenges for this literature include effective measurement of quality and integration. Quality is difficult to measure and observe in many contexts, and notably so in health care. Integration is also difficult to measure, ranging in form from full employment to complete independence with intermediate contractual relationships also possible.[4] Perhaps as a result of these challenges, stakeholders continue to vigorously debate both the sign and the magnitude of integration effects in health care markets (Gottlieb, 2013; Blankenthorn, 2009). Firms argue before the competition agencies and the courts that vertical combinations achieve integration efficiencies that would offset any anticompetitive effects. Yet despite these arguments, the Department of Justice and the Federal Trade Commission recently brought cases before the courts that alleged anticompetitive harm from combinations that would have resulted in vertical integration.[5] Consistent with the literature, the courts arrived at different conclusions regarding effects in these matters.

This paper contributes to our understanding of the effects of vertical integration in health care. We exploit novel data and a robust estimation strategy to address the challenges associated with the measurement of vertical integration and quality effects. Specifically, we measure the effects of 28 physician practice acquisitions by hospitals on a variety of health outcomes related to the treatment of hypertension and diabetes among Medicare patients. Our analysis extends the existing empirical literature on the effects of provider integration on health care quality in three ways.

First, we identify changes in integration through physician acquisitions by hospital systems. Such acquisitions sharply alter an aquired physician's status from an independent provider to a system employee. Direct employment represents an extreme form of integration that allows the new firm to manage clinical practices using financial incentive strategies that may be unavailable through looser forms of integration. Thus, by tracking changes associated with acquired physicians and comparing them to a control group, we should

Received for publication June 18, 2018. Revision accepted for publication January 17, 2020. Editor: Amitabh Chandra.

*Koch, Wendling, and Wilson: Federal Trade Commission.

We are grateful for helpful comments from Amitabh Chandra and the referee. In addition, we thank Dan Hosken, John McAdams, Luke Olson, Nathan Petek, Elizabeth Schneirov, Devesh Raval, David Schmidt, Bob Town, Christina DePasquale, Brad Shapiro, and audiences at the 2016 ASHE, 2018 ASHE, and 2018 IOS Conferences for helpful comments. Ilana Salant provided excellent research assistance. All remaining errors are our own. The views expressed in this article are those of the authors; they do not necessarily represent those of the Federal Trade Commission or any of its Commissioners.

A supplemental appendix is available online at https://doi.org/10.1162/rest_a_00924.

[1]The increase in the share of integrated systems has occurred due to both new physicians' preference for employment rather than solo practice and the integration of practicing physicians into hospital systems. Burns, Goldsmith, and Sen (2013) provide a detailed characterization of the changes in physician market structure over time. Market structure changes include greater horizontal consolidation (Rebitzer & Votruba, 2011; Capps, Dranove, & Ody, 2017). For practitioner-oriented commentary on the changes, see the Advisory Board Company's discussion of recent evidence at http://www.advisory.com/Daily-Briefing/2012/07/10.

[2]Economic theory suggests that integration effects are ambiguous and may depend on a variety of conditions including agency concerns (Cooper et al., 2005), transaction costs (Bresnahan & Levin, 2012), information asymmetries (Wolinsky, 1993; Afendulis & Kessler, 2007), and competitive incentives (Gaynor & Vogt, 2000; Whinston, 2006).

[3]For reviews, McWilliams (2013), Burns et al. (2013), and Post, Buchmueller, and Ryan (2017).

[4]Health services researchers are interested in the effects of clinical integration. Clinical integration may include full financial integration, but also other forms of provider coordination that do not necessitate formal mergers and acquisitions. See http://www.thecamdengroup.com/thought-leadership/blog/clinical-integration-an-overview/.

[5]The Department of Justice challenged the proposed merger of AT&T and Time Warner in 2018 due to vertical concerns (https://law.justia.com/cases/federal/apellate-courts/cadc/18-5214/18-5214-2019-02-26.html). The Federal Trade Commission challenged St. Luke's health system acquisition of Saltzer medical group as a horizontal merger of competitors. For more details, see http://www.ag.idaho.gov/consumerProtection/pendingActions/StLukesFTC&IdahoBrief.pdf. In both cases, the defendants argued that efficiencies would offset any anticompetitive effects.

The Review of Economics and Statistics, October 2021, 103(4): 725–739

https://doi.org/10.1162/rest_a_00924

be able to identify any effects of integration associated with our health outcomes of interest.

Second, we measure health using several direct outcome measures that represent the progression of diabetes or hypertension into worse health states: mortality, acute circulatory conditions, acute myocardial infarction (AMI), ischemic heart disease, glaucoma, and diabetes complications. Diabetes and hypertension are serious, prevalent, and treatable medical conditions. Providers track the measures we employ in order to manage patient health. In addition, medical science has correlated our outcomes with the progression of these underlying conditions.[6] Since the outcomes are health states and not treatments, they do not confound clinical effects with changes in financial incentives potentially associated with utilization-based measures, such as readmissions. This issue is important in our context since the mergers we consider likely change the financial incentives to use hospital services.

Third, we employ a variety of estimation procedures to control for potential confounding factors such as patient demographic characteristics, patient health, treatment trends, provider utilization, and provider characteristics. We use difference-in-differences and two different propensity score matching techniques to demonstrate the robustness of our results.

Our analyses find that vertical integration typically has statistically insignificant effects on health outcomes. Although some estimates suggest modest improvements in health, especially among hypertensive patients, in most cases the effects are small in economic magnitude and not robust to specification. Indeed, we interpret our null findings to be robust. For example, we do not find evidence of improved health over time, and most individual mergers are not associated with any health improvement. At best, our results suggest that vertical integration is associated with modestly better outcomes despite the fact that the literature has found it to be associated with increased expenditures (Koch, Wendling, & Wilson, 2017; Capps, Dranove, & Ody, 2015; Baker, Bundorf, & Kessler, 2014).

We contribute to the large and growing literature on the effects of integration between hospitals and physicians on costs, prices, and utilization (Neprash et al., 2015; Keating et al., 2004; Burns & Muller, 2008; Baker, Bundorf et al., 2014, 2016; Cuellar & Gertler, 2006; Ciliberto & Dranove, 2006; Capps, Dranove, & Ody, 2015). Our analysis is most closely related to papers focused on the relationship between hospital and physician integration on the quality of care, broadly defined (see Post et al., 2017, for a comprehensive review). However, some of the literature expresses concern that existing evidence has not determined the causal relationship between provider integration and quality since many papers rely on cross-sectional relationships between the variables of interest.[7] We hope to provide new clarity using rich, direct, and well-powered measures of health outcomes, together with clear measures of integration status changes.

The rest of the paper is organized as follows. Section II describes our data. Section III presents our empirical strategy, and section IV provides motivating summary and balance statistics. Section V discusses our findings. We conclude in section VI.

II. Data

Our analysis combines ambulatory and hospital claims from Medicare with provider acquisitions identified using data from SK&A and the Levin Health Care Acquisition Reports (Levin Reports).[8] We consider the period 2005 to 2012. We describe these data below.[9]

A. *Identifying the Patients of Acquired Physicians*

Our analysis considers the health outcomes of Medicare beneficiaries. We identify beneficiaries, or patients, whose doctors are acquired by a hospital system by linking together three data sources. We start with the Levin Reports, which are annual compilations of mergers and acquisitions in the health care industry. The Levin Reports identify 28 provider acquisitions that occurred between the third quarter of 2005 and the second quarter of 2010. We match the name of the group listed in the Levin Report to provider information in the SK&A data, which link physician groups to the individual doctors they employ. The SK&A data provide the National Provider Identifier (NPI) and the Unique Physician Identifier Number (UPIN) code for the members of each physician group in its data. We use those identifiers to link to physicians identified in the claims data for the 5% sample of Medicare beneficiaries from 2005 to 2012. The control group represents patients of other providers not identified as part of these acquisitions.[10]

The Levin Reports likely miss some acquisitions, such that our control group of patients contains patients whose physicians were unobservably acquired by hospitals during the sample period. Unmeasured physician acquisitions could confound our estimates if the timing of them is systematically coincidental with the timing of mergers we observe. However, we have no reason to believe that alternative measures would be coincidental with ours.

[6] For example, the 33 Accountable Care Organization quality performance measures include "Percent of beneficiaries with diabetes whose HbA1c in poor control (>9 percent)" and "Percent of beneficiaries with hypertension whose BP < 140/90." For more information, see https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html

[7] The difficulty of estimating the impact of organizational form is not specific to health care. Mullainathan and Scharfstein (2001) note that the causal effect of firm structure is unknown except for a small literature focusing carefully on identification issues (Novak & Stern, 2008; Forbes & Lederman, 2010; Kosova, Lafontaine, & Perrigot, 2013; Wilson, 2015).

[8] We do not distinguish between provider acquisitions or mergers. The distinction does not matter for the analysis or its interpretation.

[9] We provide a supplementary data description in appendix B.

[10] We do not use other measures of ownership to identify mergers. For example, we do not use the procedures identified in Neprash et al. (2015).

B. Health Conditions and Outcomes

We consider health outcomes that enable us to measure the effects of physician acquisitions on provider treatment for diabetes and hypertension. Although we consider mortality, we do not limit our analysis to it because it is relatively rare and is only weakly related to ambulatory care. The prior literature acknowledges both the benefits and concerns of mortality as an analytical measure and addresses these issues by considering mortality alongside alternative outcomes. In particular, the literature uses hospital utilization metrics, such as ER visits and readmissions, as alternatives since they are more prevalent than mortality and are likely related to health and health care costs (see Post et al., 2017).

Unfortunately, hospital utilization measures are indirect measures of health since they measure services used to treat medical conditions rather than health, directly. Often this is a desired attribute since these metrics implicitly capture both clinical health and resource intensity. However, in our setting, this attribute threatens identification since we want to isolate the clinical benefits associated with integration. Utilization-related metrics may confound financial incentives with health. For example, Medicare's provider-based billing (PBB) policies may interact with acquisitions to change the financial incentives for treating patients in either a hospital or an office (Dranove & Ody, 2016; Koch et al., 2017; Medicare Payment Advisory Commission, 2012). In order to avoid this conflation, we employ a set of outcome measures that are severe and verifiable, similar to mortality, but they are (in most cases) more prevalent and more closely related to ambulatory care, similar to hospital utilization. However, unlike hospital utilization, our outcomes are independent of related financial incentives, such as PBB. We construct our alternative metrics using the detailed data on beneficiaries' health histories contained in the Medicare data we exploit.

We consider measures related to hypertension and diabetes progression, which are treatable chronic conditions that can progress into worse health states without proper care. Table A-3 provides the descriptions for the five-digit ICD-9 codes that we use as diabetes outcomes 250.00 to 250.93. We categorize these ICD-9 codes as either "symptomatic" or "asymptomatic."[11] The codes explicitly identify conditions that are related to the progression of diabetes. For example, the description for outcomes 250.10 to 250.13 is not simply "ketoacidosis," but rather "diabetes with ketoacidosis." The description provides a relationship between the outcome and the underlying chronic condition. In addition, the existence of the code suggests that providers monitor this complication to assess the progression of diabetes. Indeed, the Agency for Healthcare Research and Quality (AHRQ) uses readmissions associated with these ICD-9 codes as quality measures in other applications.[12] The explicit relationship between the underlying chronic condition and the complication that follows is ideal for relating health outcomes to changes in treatments that arise from changes in provider ownership.

In addition to diabetes, we consider ICD-9 diagnosis codes to identify health outcomes related to hypertension progression. We consider any acute cardiac condition, which we define as any five-digit ICD-9 diagnosis code that is not defined as chronic by the National Household Interview Survey (NHIS) and is in the "conditions of the circulatory system" chapter heading of the ICD-9 code list.[13] We also identify ischemic heart disease and heart attacks, or acute myocardial infarctions (AMI), using ICD-9 codes in the claims records.[14] We use the beneficiary summary files to identify when and whether a beneficiary died.

The validity of the health outcome metrics relies on accurate, detailed, and consistent coding by the claims processors. Acquisitions resulting in strategic coding may bias our measures. For example, acquisitions that improve monitoring could appear as worsening health since they may result in more diagnoses. We address this concern by considering a battery of outcome metrics that include severe outcomes with obvious symptoms that are unlikely to be coded inconsistently across claims processors, such as mortality and acute myocardial infarction (AMI).[15] Several of our outcome measures have been used extensively by the previous literature in different contexts for similar reasons.

We treat all of our health conditions as absorbing states for the beneficiary. Individuals observed with a health outcome are defined as having the condition for the rest of the sample period regardless of whether the contemporaneous period contains a claim with the ICD-9 diagnosis listed. We impose this restriction for all health outcomes, including acute conditions. In addition, we omit 2005 data from our analysis, limiting the analysis to the period 2006 to 2012. We do this to provide an entire year of observed claims to determine the health conditions and (potential) outcomes of beneficiaries observed in 2006.[16]

III. Econometric Approach

When using nonexperimental data to identify the impact of an event, it is important to control for factors associated with both the event and the outcome of interest. For example, one might be concerned about compositional differences in the characteristics of providers and patients across our treatment

[11] The categories are based on the ICD-9 code descriptions. "Symptomatic" complication descriptions are likely experienced by a patient through symptoms (e.g., blindness). "Asymptomatic" condition descriptions may be unknown to a patient who is not monitoring her body chemistry, such as blood sugar levels.

[12] https://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V21/pqi_guide_rev4.pdf

[13] The relevant ICD-9 codes in the chapter covering circulatory conditions range from 390 to 459. We replicate the NHIS table of chronic conditions as table A-8. We define AMI as the three-digit ICD-9 code 410, ischemic heart disease as either 411 or 414, and glaucoma as 365.

[14] We find similar qualitative results when we consider the annual incidence of these outcomes provided in the beneficiary summary files.

[15] Our symptomatic diabetes metric also has this property.

[16] We exclude beneficiaries who visit a merged physician during 2005, and only during 2005.

and control samples. To control for compositional differences between the samples, we identify the effect of integration through changes in the integration status of several physician groups. However, this difference-in-differences strategy relies on assumptions relating the treated and control group samples, such as parallel trends. These assumptions may be violated if, for example, hospitals target growing or innovative practices for acquisitions. To address this concern, we also employ two different propensity score estimators to identify a set of relevant control-group patients in the context of difference-in-differences. Specifically, we use a propensity score stratification estimator and a single nearest neighbor propensity score match without replacement within an exact match (Rosenbaum & Rubin, 1983; Imbens & Rubin, 2015).

Propensity score techniques are useful for measuring treatment effects if covariate distributions differ substantially by treatment status.[17] If so, parameter estimates may be subject to confoundedness bias. Although these concerns are especially pronounced in repeated cross sections, we are able to address this issue in the context of difference-in-differences (Abadie, 2005). We demonstrate that matching reduces the potential for confoundedness in the full sample. The matched sample also allows us to show that the parallel trends assumption holds for this sample.[18] Moreover, matching allows us to flexibly control for differences between the groups. However, matching substantially reduces the size of the sample. Blocking is an alternative procedure that helps reduce potential confoundedness, similar to matching, but employs the full sample.

A. *Propensity Score Matching and Stratification*

Our sample is a good candidate for using propensity score procedures. We have a large set of potential control-group patients relative to the set of patients of acquired physicians, and the full sample differs from the treatment sample with respect to observable characteristics. We estimate the propensity score using a logistic probability model that relates acquisition characteristics to provider characteristics, d; group characteristics, g; and patient characteristics, i, associated with the beneficiaries. The following person-level equation represents the probability equation that we use to match patients:

$$Pr(A_i = 1|X; \Theta_A) = f(\alpha + \beta_i X_i^i + \beta_d X_i^d + \beta_g X_i^g + \epsilon_i), \quad (1)$$

where X_i^i are patient characteristics for patient i, X_i^d are the characteristics of providers visited by patient i, and X_i^g are characteristics of the physician groups visited by patient i.

We account for patient heterogeneity, X_i^i, using the patient's birth cohort, race, sex, health condition, and urbanicity.[19] Patient health is measured using indicators for whether a patient had a chronic condition within each of the major ICD-9 chapter heading categories and indicators for whether the patient had a mental health condition, an injury, or an infectious disease.[20] We measure urbanicity using the patient's county of residence matched to the 2009 and 2015 Area Health Resource Files (AHRF).[21] We associate patient demographic conditions that change during the sample, such as urbanicity and population density, with the values observed at the beginning of the sample.[22] However, we associate a patient with health conditions observed at the end of the period to ensure that patient matches are treated for the same conditions observed in our sample of acquired patients.

Provider characteristics, X_i^d, include the number of physicians visited and 21 discrete specialty categories of the physician.[23] For patients who visit multiple physicians, we associate the patient with all of the relevant criteria such that each of the categorical variables is not mutually exclusive.[24]

Finally, we control for variation in physician group characteristics, X_i^g, using the size of the practice as measured by the number of providers with the same TAXID.[25] Again, we associate beneficiaries who visit multiple physicians with all of the relevant practice variables for those physicians.[26]

We construct the propensity score for each patient using the predicted probability estimates from equation (1). We use the propensity score to construct a matched sample of patients of acquired and nonacquired physicians. We describe our matching procedure in appendix C. The matched sample is a quarterly data set that follows each pair of treated and matched patients over time. We also use the propensity score to divide the full sample into strata, or blocks, for our blocking

[17] Henceforth, we use acquisition effects to refer to the treatment effects discussed in the econometrics literature. We distinguish acquisition effects from treatment effects since clinical treatments have a role in our discussion of the results, and thus create a potential source of confusion.

[18] See the parallel trend graphs reported in appendix section III.1.

[19] The logistic equation is estimated separately for men and women beneficiaries. State is determined by the location of the claim as defined by Medicare. State is interacted with urbanicity. We also control for population density. Birth cohort includes a linear "age in 2005" term and its square as well as seven bins for under 70, 70–74, 75–79, 80–84, 85–89, 90–94, and over 94 in 2005. Race is defined as non-Hispanic white, black, and other.

[20] We produce the ICD-9 chapter headings as table A-1. Mental health conditions, injuries, and infectious diseases are rarely defined as chronic conditions.

[21] Urbanicity takes on four values: a county in a metro area with more than 1 million people, a metro area with 500,000 to 1 million people, a metro area with 250,000 to 500,000 people, and a nonmetro area. We also control for the natural logarithm and its square of population density.

[22] Patients who enter at the end of the sample are matched to the later resource file.

[23] We categorize three specialties as family medicine.

[24] For example, if a beneficiary visits two different cardiologists, the patient is associated with a single dummy variable representing cardiology. However, if a beneficiary visits a cardiologist and a gastroenerologist, we associate the patient with separate dummy variables for each specialty.

[25] We use the minimum firm size observed during the period and bin the sizes into five categories: fewer than 5, 5 to 24, 25 to 49, 50 to 99, 100 to 199, and over 200 and above. We choose the minimum since it likely better reflects the size of the acquired group at the time of the acquisition.

[26] For example, if a beneficiary visits two small practices (i.e., fewer than five physicians), then we associate the beneficiary with a small physician practice. If, however, the beneficiary visits two physicians, one at a small practice and another at a large practice, we associate the beneficiary with both the large and the small practice. Both the large and the small practice are defined as such using the minimum number of providers at the firm during the sample period.

estimator. We describe how we construct the blocks and how they are used in estimation in appendix D.

B. Health Outcome Acquisition Effects

Our primary specification estimates health outcome acquisition effects using a discrete-time hazard model. The model controls for a rich set of patient demographic characteristics, provider characteristics, and quarter-year dummies during the period 2006 to 2012.[27] We represent this specification with the following equation:

$$Pr(h_{it} = 1|X_{it}, \Theta) = \Lambda(\alpha + \theta_{PM}PM_{it} + \beta_M M_i + \delta_{ZIP} + \delta_t + \beta_x X_{it} + \beta_{M*y}(M_i \times yr_{it}) + u_{it}). \quad (2)$$

This model considers the probability that we observe a positive realization of our health outcome variable in the contemporaneous period, $h_{it} = 1$. We separately model each health outcome (acute cardiac conditions, AMI, death, diabetes complications, glaucoma, and ischemic heart disease) as a discrete indicator that specifies whether the patient is observed with the outcome during the quarter, or not. As is standard for hazard models, we model each outcome as an absorbing state.

The postmerger coefficient, θ_{PM}, is the effect of interest in equation (2). The postmerger variable, PM, represents the interaction of the acquisition indicator, M, with an indicator that is equal to 1 if the contemporaneous quarter follows the earliest relevant acquisition.[28] We interpret the coefficients as the clinical benefits (or harm) attributable to the acquisition. For our matched sample, we interact a female indicator with the merger coefficients and age to allow for differential effects across men and women.[29] Blocking and OLS are estimated separately for men and women.

The model also controls for patient and provider characteristics, X_{it}, quarter-year fixed effects, δ_t, and three-digit ZIP code fixed effects, δ_{ZIP}, which are similar to a county but are geographically smaller in denser areas and thus better account for population.[30] The demographic controls include ICD-9 major category chronic condition indicators and age fixed effects.[31] The specification includes separate indicators for each age up to 95 and an indicator for whether the patient is older than 95. For most of our health outcomes, we limit the sample to patients with diabetes or hypertension, as identified using the ICD-9 indicators in the contemporaneous quarter.[32] The provider controls include indicators for the seven most prevalent physician specialties in our sample, including family medicine, as well as an indicator for whether the beneficiary visited another specialist not included on the list of the most frequently observed specialties in our data.[33] We also allow for an acquisition-specific time trend, β_{M*y}, to differ from the jointly estimated quarter-year fixed effects.[34]

The full sample estimators (blocking and OLS) assume that $\Lambda(\cdot)$ takes a linear functional form, whereas all of our results using the matched sample estimator assume that $\Lambda(\cdot)$ takes the form of a proportional hazard model.[35] The matching and OLS estimators simply estimate the acquisition effects among their respective samples. The matching estimator clusters the standard errors by the matched group (i.e., the treated patient and her matched pair) to allow for correlations over time and across birth cohort-sex-specialty-propensity score matches. Since the OLS estimator does not have matching treatment and control pairs, we cluster by the three-digit ZIP code of the beneficiary. The blocking estimator estimates separate acquisition effects for each block, or strata, within the full sample and constructs the acquisition effect from a weighted average of the within-block estimates, assuming independence across blocks. We describe this aggregation procedure in appendix D. The blocking estimator also clusters standard erros by the three-digit ZIP code within a block.

C. Robustness to Match-Specific Unobservables

In order to control for unobservable differences at the matched-pair level and fully take advantage of the matching procedure, we estimate a fixed-effects logit model via the Chamberlain (1980) procedure.[36] Unfortunately, this procedure does not allow us to recover the marginal effects of the estimator. In order to interpret the estimates from the fixed-effects matching estimator, we compare the resultant coefficient estimates from this analysis to that of the more parsimonious specification, equation (2). The comparison, reported in table 4, enables us to determine whether controlling for matched-pair fixed effects alters our conclusions relative to the more parsimonious equation. As in our other matching

[27] The fixed-effect proportional hazard includes quarter and year dummies but not their interaction.

[28] We use the closing dates to determine acquisition timing when available and announcement dates otherwise. The omitted period is the observed quarter prior to the acquisition. The postacquisition interaction for patients of nonacquired physicians is always 0 and thus not included.

[29] The matched sample includes a control for sex, interactions of sex with all of the merger variables, and interactions of sex with the following age categories: 65–69, 70–74, 75–79, 80–84, 85–89, 90–94, and 95 and above.

[30] In each of the matching specifications, we pool ZIP codes with fewer than 1,250 observations and interact the resultant dummy with the state of residence.

[31] We also include dummies for mental health, injuries, and infectious health conditions.

[32] We limit the sample to patients with hypertension when we consider the acute heart conditions, ischemic heart disease, and AMI health outcomes. We limit the sample to patients with diabetes when we consider the two diabetes complications and glaucoma. We consider mortality among the entire sample population.

[33] We define family medicine doctors as specialists in general practice, family practice, or internal medicine. We also define visits to physician assistants as family medicine.

[34] This parameter is identified separate from the postmerger effect since the acquisitions occur at different times during the sample.

[35] This includes the specifications, described below, that interact the acquisition coefficients with specific mergers and lead and lag acquisition timing variables.

[36] Appendix III.2 discusses this specification and its results. This specification includes quarter and year dummies rather than interactions of quarter and year.

estimator specifications, we cluster the standard errors by matched pair.

D. Robustness to Heterogeneous Acquisition Effects

To address the possibility that our average treatment estimates mask heterogeneity, we estimate two specifications that allow for differential merger effects across time and acquisition. First, we estimate separate effects for each acquisition to allow for the possibility that acquisitions have differential effects. We represent this specification with equation (3):

$$\begin{aligned} &Pr(h_{it} = 1|X_{it}, \Theta) \\ &= \Lambda\Bigg(\alpha + \sum_{g\in G} \theta_p^g M_{it}^g Post_{it}^g + \sum_{g\in G} \theta_M^g M_{it}^g + \beta_x X_{it} \\ &\quad + \beta_{M*y}(M_i \times yr_{it}) + \delta_{ZIP} + \delta_t + u_{it}\Bigg). \end{aligned} \tag{3}$$

Equation (3) allows the postacquisition effect to vary by each of the G acquired groups. The specification allows for both a different practice effect, θ_M^g, and a different postmerger effect, θ_p^g. However, we pool smaller acquisitions together and estimate a single effect for them. We also do not estimate separate acquisition effects for men and women, although we continue to control for separate age effects for men and women.

Second, we estimate a specification that allows the acquisition effect to depend on the time since the acquisition occurred, including periods prior to consummation:

$$\begin{aligned} &Pr(h_{it} = 1|X_{it}, \Theta) \\ &= \Lambda\Bigg(\alpha + \sum_{l=-15}^{15} \theta_l l_{it} + \sum_{l=-15}^{15} \theta_{fl}(l_{it} \times fem_i) + \beta_M M_i \\ &\quad + \beta_{fM}(fem_i \times M_i) + \beta_x X_{it} + \beta_{M*y}(M_i \times yr_{it}) \\ &\quad + \delta_{ZIP} + \delta_t + u_{it}\Bigg). \end{aligned} \tag{4}$$

Equation (4) is nearly identical to equation (2), except that we allow the effect of the acquisition to vary by the time since consummation, l_{it}. Each l_{it} is a dummy variable that is equal to 1 if the contemporaneous quarter was l months since the earliest observed acquisition and 0 otherwise. We omit the period of the acquisition and set all acquisition variables to 0 if the relevant group did not experience an acquisition during our sample period. We cautiously use the quarter-specific postmerger coefficient estimates to determine whether merger benefits take time to become realized.

IV. Descriptive and Motivating Statistics

In table 1, we provide summary statistics characterizing a select set of important characteristics for our Medicare sample. Each observation represents a patient-quarter combination. We manually fillin data for quarters where patients do not have claims. We assume that health states do not change during the intervening periods within a year since a change in health status would likely initiate a medical claim.[37] Table 1 provides all information separately for patients of acquired physicians, patients of potential control group physicians (i.e., all nonacquired providers), and the sample of acquired patients and their matches using our matching methodology.[38]

For each subsample, table 1 provides the average demographic characteristics such as age, sex, race, urbanicity, and health condition of the beneficiaries. Table 1 also provides the average propensity score of the patients and characteristics of the providers seen by the beneficiaries, including the specialties of the providers and the firm size of visited providers.[39] Beneficiaries in the sample of potential controls are, on average, 76.7 years old, 87.1% white, 37.0% male, and live predominantly in metro areas with more than 500,000 people. Beneficiaries often visit multiple providers during a quarter: 85.4% of potential control beneficiaries visit a provider in a group with fewer than 5 providers, and 68.8% visit physicians employed by groups with between 5 and 24 providers. The vast majority of beneficiaries have some health condition, as 91.9% of potential controls have at least a chronic circulatory condition.

Table 1 reports normalized differences in order to provide a metric of the compositional differences between the samples for each of the variables that we report. The normalized differences are a balance statistic, a standard measure reported in the propensity score literature used to evaluate the performance of the matching methodology (see Imbens & Wooldridge, 2009).

The normalized differences column labeled "Match" in table 1 compares the sample of patients visiting acquired physicians against the sample of patients matched using the propensity score matching procedure described in appendix C. Normalized differences closer to 0 suggest more sample balance, and thus a better control group, than normalized differences that are different from 0. Rosenbaum and Rubin (1985) suggest that a normalized difference with absolute value of less than 0.2 likely provides good balance between the sample and control group, although the 0.2 threshold is a rule of thumb.

The normalized differences suggest that the full sample of potential controls and the sample of patients visiting acquired physicians differ substantially from each other with respect to provider characteristics and the propensity score.

[37] Beneficiaries in our sample visit a provider at least once during the year. We fillin between 9% (diabetes sample) and 14% (full sample) of quarterly information. Death is obtained from beneficiary summary files rather than claims information.

[38] We provide the list of acquisitions and their states in appendix table A-4.

[39] Patients visit multiple physicians during a quarter, such that the physician specialties and firm size variables are not mutually exclusive.

TABLE 1.—SUMMARY AND BALANCE STATISTICS FOR THE MEDICARE SAMPLE, 2006–2012

	All potential controls		Matched nonacquired		Acquired		Normalized differences	
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Match	Potential Controls
Propensity Score	0.023	0.048	0.136	0.117	0.145	0.126	0.074	1.274
				Demographic characteristics				
White	0.871	0.336	0.918	0.275	0.917	0.276	−0.003	0.151
Male	0.370	0.483	0.380	0.485	0.383	0.486	0.005	0.026
Age	76.71	7.81	77.66	7.44	77.60	7.44	−0.008	0.117
Metro >1MM	0.450	0.497	0.468	0.499	0.466	0.499	−0.004	0.032
Metro 500K-1MM	0.211	0.408	0.248	0.432	0.230	0.421	−0.042	0.045
Metro <500K	0.115	0.319	0.093	0.291	0.096	0.295	0.010	−0.061
Nonmetro	0.223	0.417	0.191	0.393	0.208	0.406	0.043	−0.038
				Health condition diagnoses				
Hypertension	0.863	0.344	0.900	0.300	0.905	0.294	0.015	0.131
Diabetes	0.374	0.484	0.379	0.485	0.384	0.486	0.010	0.020
Circulatory	0.919	0.272	0.953	0.213	0.955	0.206	0.014	0.150
Musculoskeletal	0.817	0.387	0.868	0.339	0.872	0.335	0.012	0.151
Sense organs	0.669	0.470	0.725	0.447	0.723	0.447	−0.003	0.117
Gastrointestinal	0.636	0.481	0.709	0.454	0.713	0.452	0.011	0.166
Injury	0.597	0.491	0.667	0.471	0.675	0.469	0.017	0.162
Endocrine	0.584	0.493	0.600	0.490	0.602	0.490	0.004	0.037
Signs/Symtoms	0.555	0.497	0.630	0.483	0.645	0.479	0.029	0.183
Respiratory	0.525	0.499	0.591	0.492	0.599	0.490	0.016	0.149
Skin conditions	0.440	0.496	0.484	0.500	0.490	0.500	0.012	0.101
				Provider characteristics				
Family practice	0.560	0.496	0.625	0.484	0.634	0.482	0.020	0.152
Other	0.429	0.495	0.480	0.500	0.483	0.500	0.005	0.107
Radiology	0.209	0.407	0.251	0.434	0.259	0.438	0.016	0.117
Cardiology	0.154	0.361	0.195	0.396	0.216	0.412	0.053	0.162
Opthalmology	0.129	0.335	0.141	0.348	0.143	0.350	0.004	0.041
Podiatry	0.100	0.300	0.118	0.323	0.117	0.322	−0.002	0.055
Firm size <5	0.854	0.353	0.850	0.357	0.848	0.359	−0.004	−0.017
Firm size 5–24	0.688	0.463	0.750	0.433	0.755	0.430	0.012	0.150
Firm size 25–49	0.376	0.484	0.453	0.498	0.501	0.500	0.096	0.254
Firm size 50–99	0.320	0.467	0.407	0.491	0.457	0.498	0.100	0.283
Firm size 100–200	0.296	0.457	0.387	0.487	0.419	0.493	0.065	0.258
Firm size >200	0.323	0.468	0.491	0.500	0.491	0.500	−0.001	0.346
Observations	39,826,077		1,093,109		1,030,742			

The normalized differences are greater than 0.2 with respect to several of the firm size variables we consider. The acquisition sample is more likely to visit a physician in a large firm than the sample of potential controls. In addition, the normalized differences of the propensity score estimates are greater than 1.2.

The matching procedure is effective in reducing the normalized differences between samples for every variable we consider. The normalized differences between the patients of acquired physicians and the matched sample are always less than or equal to 0.1. Thus, the matching procedure results in a matched sample that is well balanced against the acquired sample for all of the variables that we consider, including the provider characteristic variables that resulted in sample imbalance between the acquired sample and the full sample of potential controls. The matching procedure also results in greater balance for all of the other demographic variables such as age, sex, race, diagnosed health, and provider specialty, for which the potential control sample was relatively similar to the acquisition sample. Appendix tables A-5 and A-6 provide similar summary and balance statistics as table 1 but among diabetic and hypertensive beneficiaries, respectively.[40] These comparison patterns are similar to those found when we do not condition on hypertension or diabetes.

Next, we consider the average health outcomes that we use as the dependent variable in our analyses. The health outcomes that we consider are indicators for whether the beneficiary transitioned into the health condition during the quarter. Table 2 provides the mean outcome variables for patients of acquired physicians prior to the acquisition, patients of acquired physicians after the acquisition, and the matched patients we use as a control group for the patients of acquired physicians.

A comparison of outcomes among patients of acquired physicians in the preacquisition period against outcomes of matched beneficiaries finds that average outcomes are somewhat better among the matched beneficiaries. For example, among women beneficiaries with hypertension, 4.27% of patients in the matched sample develop acute cardiac

[40]Note that diabetes is a chronic endocrine condition, and hypertension is a chronic circulatory condition.

TABLE 2.—SUMMARY OF NEW CONDITION DIAGNOSES FOR MATCHED SAMPLE, 2006–2012

	Matched Sample			Preacquisition			Postacquisition		
Outcome	Observations	Mean	Standard Deviation	Observations	Mean	Standard Deviation	vations	Mean	Standard Deviation
Women				Full sample					
Mortality	677,829	0.0123	0.1104	393,075	0.0107	0.1028	243,334	0.0157	0.1243
				Diabetes sample					
Mortality	244,369	0.0171	0.1297	133,231	0.0151	0.122	99,570	0.0198	0.1393
Asymptomatic	223,094	0.0058	0.0763	121,523	0.008	0.089	88,168	0.0052	0.072
Glaucoma	181,275	0.0055	0.0743	102,918	0.0056	0.0745	71,285	0.0044	0.0665
Symptomatic	155,016	0.0259	0.1589	85,174	0.0317	0.1752	57,836	0.0214	0.1447
				Hypertension sample					
Mortality	613,928	0.0132	0.1141	349,180	0.0115	0.1068	230,092	0.0163	0.1267
AMI	574,484	0.0058	0.076	328,527	0.0070	0.0835	207,225	0.0056	0.0748
Acute cardiac	323,628	0.0427	0.2021	203,296	0.0455	0.2084	95,219	0.0396	0.195
Ischemic heart	361,774	0.0250	0.1560	217,080	0.0286	0.1667	117,934	0.0184	0.1343
Men				Full sample					
Mortality	415,280	0.0138	0.1166	247,730	0.0123	0.1101	146,603	0.0168	0.1286
				Diabetes sample					
Mortality	169,387	0.018	0.133	95,730	0.0165	0.1275	66,790	0.0208	0.1428
Asymptomatic	153,737	0.0064	0.08	88,905	0.0069	0.0829	60,221	0.0057	0.075
Glaucoma	131,769	0.0052	0.0718	78,735	0.0057	0.0751	51,177	0.0043	0.0651
Symptomatic	106,175	0.028	0.1649	62,009	0.0344	0.1822	38,215	0.0233	0.1507
				Hypertension sample					
Mortality	370,265	0.0148	0.1207	215,780	0.0134	0.1148	137,538	0.0175	0.1311
AMI	335,366	0.0079	0.0887	198,270	0.0093	0.0959	119,150	0.0074	0.0855
Acute cardiac	189,629	0.0458	0.209	121,420	0.0503	0.2185	55,094	0.0411	0.1984
Ischemic heart	157,908	0.0359	0.186	94,964	0.0435	0.2039	45,535	0.0257	0.1583

conditions, whereas 4.55% of patients in the preacquisition period develop acute cardiac conditions. These patterns appear for most of the outcomes that we consider. Mortality is an important exception in most samples. However, the comparison of averages does not control for other factors that may explain the differences, such as patient demographics, provider characteristics, industry trends, or length of the sample.[41]

Average health typically improves following an acquisition among patients of acquired physicians. For example, among women with hypertension, 4.55% of acquired patients develop an acute cardiac condition in the preacquisition period, and 3.96% develop such conditions in the postacquisition period, a better outcome, on average. However, we find that mortality is worse, on average, in the preperiod than in the postperiod. For example, women face 1.07% mortality in the preperiod but a 1.57% mortality rate in the postmerger period. However, worse outcomes may simply represent aging of the population, and better outcomes may represent technological improvement in the health care system. The goal of our analysis in the following sections is to determine whether the mean differences observed in table 2 survive the inclusion of controls for potentially confounding factors such as age and general trends in health and health care. If so, the differences may represent acquisition effects.

[41] The sample length may be an important consideration in this comparison since patients in the preacquisition period are observed for a much shorter period than are their matched counterparts.

V. Empirical Results

A. Main Results

In this section, we present our estimates derived from estimation of equation (2). Table 3 presents marginal effects and their respective standard errors using three types of estimators. The "Difference-in-Difference" columns represent postmerger coefficients from a linear probability model estimated using the full sample. The "Blocking Estimator" columns represent the weighted average of postmerger coefficients from a linear probability model, estimated separately within blocks defined using the propensity score. "Full Control Matching" represents the marginal effects from a proportional hazard model using pairs of patients of acquired physicians and their match.[42]

We split table 3 into three panels from top to bottom and report all results separately for men and women. The top panel considers the effects of acquisitions on mortality for the full sample, not limited by condition. The middle panel

[42] All results are reported separately for men and women and are derived from models that control for age, gender, race, quarter fixed effects, firm size, number of physicians seen, the beneficiary's ZIP code, physician specialty, an acquisition-specific time trend, major chapter chronic condition indicators and indicators for injuries, mental health conditions, and infectious diseases. The difference-in-difference and blocking estimators estimate separate models for men and women and control for the five-digit ZIP code of the patient interacted with race. The matching estimator controls for the three-digit ZIP code of the patient and interacts age, race, and the effect of interest with sex.

TABLE 3.—POSTACQUISITION EFFECTS ON HEALTH-STATE TRANSITION PROBABILITIES FROM THREE ESTIMATORS

	Difference-in-difference		Blocking estimator		Matching estimator	
Outcome	Women	Men	Women	Men	Women	Men
			Full sample			
Mortality	0.0012*	0.0008	0.0018	0.0020	0.0130	0.0003
	(0.0004)	(0.0005)	(0.0031)	(0.0073)	(0.0093)	(0.0086)
			Diabetes sample			
Mortality	0.0013*	0.0000	0.0022	0.0021	−0.0061	−0.0001
	(0.0006)	(0.0009)	(0.0051)	(0.0064)	(0.0132)	(0.0122)
Asymptomatic	−0.0007	0.0004	−0.0004	0.0010	−0.0037	0.0216
	(0.0004)	(0.0005)	(0.0036)	(0.0043)	(0.0188)	(0.0209)
Glaucoma	0.0004	0.0002	0.0003	0.0005	−0.0053	−0.0108
	(0.0005)	(0.0006)	(0.0037)	(0.0045)	(0.0254)	(0.0273)
Symptomatic	−0.0028*	−0.0030	−0.0022	−0.0017	0.0113	0.0083
	(0.0012)	(0.0015)	(0.0082)	(0.0118)	(0.0128)	(0.0132)
			Hypertension sample			
Mortality	0.0011*	0.0008	0.002	0.0022	0.0141	−0.0005
	(0.0004)	(0.0005)	(0.0035)	(0.0071)	(0.0094)	(0.0087)
AMI	−0.0013*	−0.0019*	−0.0011	−0.0018	−0.0359*	−0.0097
	(0.0003)	(0.0005)	(0.0024)	(0.0066)	(0.013)	(0.0118)
Acute cardiac	−0.003*	−0.0059*	−0.0019	−0.0061	0.0079	0.0020
	(0.001)	(0.0015)	(0.0087)	(0.0193)	(0.0071)	(0.0070)
Ischemic heart	−0.0029*	−0.0077*	−0.0014	−0.0013	−0.0183	−0.0214*
	(0.0007)	(0.0016)	(0.0061)	(0.0213)	(0.0094)	(0.0107)

*Statistically significant at the 5% C.I. Standard error in parentheses. Marginal effects are percentages/100. Marginal effect estimates are calculated as $ME = \Lambda(X\hat{\beta} + \hat{\theta}_{PM}) - \Lambda(X\hat{\beta})$ for a 77-year old in the 452 three-digit ZIP code among the treated 1Q2006 for matching. Controls include race, physician specialty, patient ZIP code, major ICD-9 condition characteristics, age, firm size, number of doctors seen, age category interactions with sex, and quarter dummies. OLS and blocking also control for five-digit ZIP codes interacted with race, are estimated separately for men and women, and control for more specialties. Linear model estimates are simply the postmerger coefficient estimates. The male difference-in-difference mortality estimate is 3.8×10^{-6}.

considers the effects of acquisitions on mortality, glaucoma, asymptomatic diabetes complications, and symptomatic diabetes complications among diabetic patients. The third panel considers the effects of acquisitions on mortality, AMIs, acute cardiac conditions, and ischemic heart disease among hypertensives. The results reflect percentage effects divided by 100.[43] For example, we interpret the 0.0003 estimate for men among the full sample to imply that integration increases the probability that men will die by 0.03%.[44] However, the marginal effect estimate is statistically insignificant at the 5% confidence level, which is consistent with no merger effect.

Of the 54 marginal effect estimates reported in table 3, we find that 12 are statistically significant at the 5% confidence interval, but 10 of those 12 are from our difference-in-difference estimator, which may suffer from bias due to potential confoundedness reported in table 1. The blocking estimator does not find statistically significant effects for any of the outcomes that we consider. The matching estimator does not not find statistically significant effects of acquisitions for diabetes outcomes or mortality.

Most of the statistically significant results are among the hypertensive sample. Both the matching estimator and the OLS difference-in-difference estimator find that acquisitions reduce progression into ischemic heart conditions and heart attacks. Perhaps these results provide some evidence that integration improves cardiac health outcomes. However, the OLS difference-in-difference estimator results are economically small.[45] In addition, the matching estimator does not find that acquisitions improve any outcome for both sexes, and the blocking estimator does not find any of the effects to be statistically significant. Finally, three of the twelve statistically significant estimates suggest that mergers increase mortality for women. However, these results are, again, from our difference-in-difference estimator and are economically small. In addition, mortality effects are not found for men in any sample for any of the estimators. Taken together, we interpret the results to suggest that mergers have little effect on the outcomes that we consider.

We draw much in our conclusions from our preferred estimator, the matching specification estimator. We view the other estimators as robustness checks against it. However, one might be concerned that our "full set of controls" matching specification estimator is not flexible enough to address differences between patients of acquired and nonacquired physicians. To address this concern, we estimate a proportional hazard model with matched pair fixed effects. We describe the procedure in appendix section III.2 and report the

[43]The reported marginal effect estimates effectively represent the change in the hazard rate attributed to the postacquisition effect for each of our health outcomes, which we calculate as $ME = \Lambda(X\hat{\beta} + \hat{\theta}_{PM}) - \Lambda(X\hat{\beta})$. Marginal effect estimates are evaluated for a 77-year-old, nonwhite patient of a merged provider who did not visit a specialist or have a chronic condition in the 452 three-digit ZIP code (Cincinnati, Ohio) during 1Q2006.

[44]We interpret the coefficient to mean that the merger increases the probability of dying since the coefficient is positive. Negative coefficients represent decreases in probabilities and thus better health outcomes.

[45]The precision of these estimates likely results from use of between 4 and 26 million observations despite clustering at the three-digit ZIP code level. Dependent variable means and regression statistics are provided in appendix table A-7 for the difference-in-difference samples.

TABLE 4.—POSTACQUISITION COEFFICIENT ESTIMATES FROM THREE MATCHING ESTIMATOR SPECIFICATIONS

	Pair fixed effects			Full set of controls			Age and quarter only		
Outcome	Women	Men	R^2	Women	Men	R^2	Women	Men	R^2
				Full sample					
Mortality	0.0869	−0.0085	0.2946	0.0535	0.0013	0.1911	0.038	−0.0142	0.0526
	(0.0487)	(0.0559)		(0.0384)	(0.0414)		(0.0351)	(0.0386)	
				Diabetes sample					
Mortality	0.0936	−0.1083	0.3588	−0.0253	−0.0004	0.1769	−0.0347	−0.0246	0.0434
	(0.089)	(0.0995)		(0.0545)	(0.0575)		(0.0496)	(0.0538)	
Asymptomatic	−0.0456	0.2844	0.215	−0.0165	0.0968	0.0792	−0.1051	−0.0015	0.0065
	(0.1244)	(0.1512)		(0.0847)	(0.0946)		(0.0792)	(0.0893)	
Glaucoma	0.1152	−0.4553	0.4206	−0.0223	−0.0445	0.0907	0.0364	−0.0387	0.0103
	(0.1949)	(0.228)		(0.1069)	(0.1132)		(0.0981)	(0.1055)	
Symptomatic	0.0774	−0.082	0.2319	0.0463	0.0355	0.0766	0.0621	0.0207	0.0106
	(0.085)	(0.095)		(0.0528)	(0.0567)		(0.0487)	(0.0529)	
				Hypertension sample					
Mortality	0.0825	−0.0224	0.2902	0.0579	−0.0025	0.1851	0.0532	−0.0164	0.0492
	(0.0501)	(0.0577)		(0.039)	(0.0421)		(0.0358)	(0.0394)	
AMI	−0.1229	−0.1465	0.3956	−0.1581*	−0.0481	0.2266	−0.1615*	−0.1000	0.0068
	(0.0744)	(0.0808)		(0.0557)	(0.0585)		(0.0504)	(0.0535)	
Acute cardiac	0.0251	−0.026	0.2936	0.0344	0.0097	0.1586	−0.0008	−0.0318	0.0076
	(0.0421)	(0.0524)		(0.031)	(0.0346)		(0.0273)	(0.0309)	
Ischemic heart	−0.1086*	−0.0166	0.3179	−0.0747	−0.087*	0.1956	−0.1365*	−0.1282*	0.0128
	(0.0538)	(0.0692)		(0.0388)	(0.0434)		(0.034)	(0.0384)	

*Statistically significant at the 5% confidence interval. Standard errors reported in parentheses. Matched-pair fixed-effects model estimated using procedure outlined by Chamberlain (1980). Fixed effects in that model represent indicators for matches between an acquired physician's patient and a nonacquired physician's patient from the propensity score procedure. Fixed-effect model includes quarter and year dummies but not the interactions. Marginal effects from the fixed-effects model cannot be recovered. Marginal effects from the other specifications provided in table C-9.

resultant coefficient estimates in table 4, alongside results from the fully specified model. Unfortunately, our fixed-effects estimation procedure does not allow us to recover marginal effects. We therefore compare coefficient estimates across specifications. The comparison shows that the fixed-effects coefficients and the full control model specifications typically reach the same inference and are similar to each other in magnitude. Indeed, the full control model finds two of the estimated eighteen coefficients are statistically significant, whereas the fixed-effects model finds that integration improves health in only one instance: hypertensive women progressing into ischemic heart disease.

Alternatively, one may be concerned that our matching specification estimator is overspecified and suffers from attenuation bias. To address this concern, we estimate a more parsimonious specification using our matching sample. We report the estimated coefficients from this specification in table 4, alongside the coefficient estimates from our full specification and our fixed-effects specification.[46] The columns "Age and Quarter Only" present the coefficient estimates from this parsimonious specification that includes controls for the patient's age and quarter-year dummies but no other controls. We find limited evidence that attenuation bias is responsible for our findings. Simply controlling for time and age is enough to conclude that fifteen of the eighteen marginal effect estimates are statistically insignificant at the 5% confidence level among our matched sample. The statistically significant coefficients suggest that these acquisitions reduce the probability that beneficiaries with hypertension have worse outcomes. However, these findings are similar to those in our fully specified model, which also finds statistically significant effects in two of the three cases.

Due to the similarity of conclusions from our preferred "full set of controls" specifications with other specifications, the following sections consider the effects of integration separately by merger and over time following an acquisition using that specification.

B. Effects Separately by Acquisition

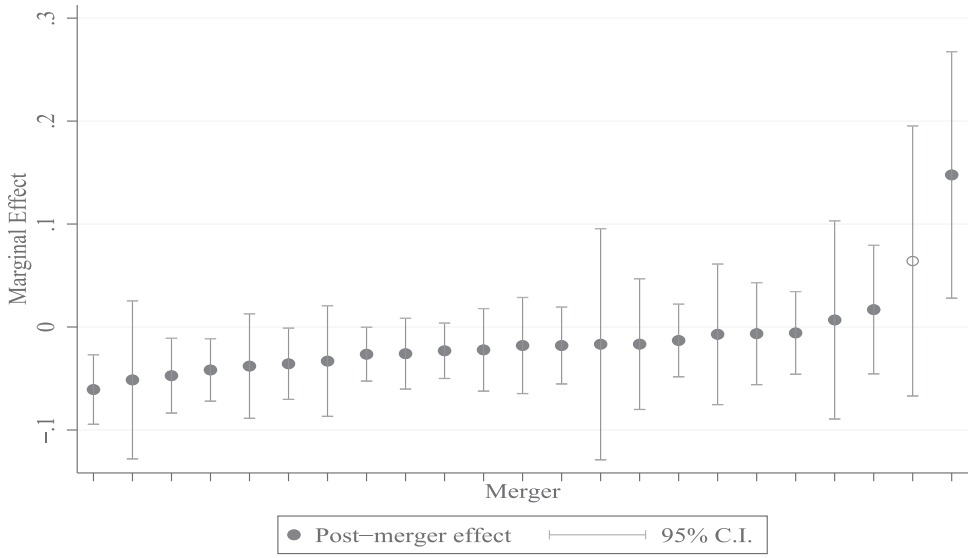
Next, we consider whether different acquisitions had different impacts on patients' health. This might occur if some firms are better at integration than others.[47] We plot our acquisition-specific marginal effect estimates and the corresponding 5% confidence intervals in figures 1, 2, and 3. In each figure, the marginal effect estimates are sorted such that the smallest marginal effect (i.e., the largest negative estimate) is plotted on the left-most part of the figure and the largest marginal effect estimate is on the right-most part of the figure.

Figure 1 considers the acquisition-specific effects of acquisitions on mortality for the full sample. The signs of all but four of our merger-specific estimates are negative. However, all but three are statistically insignificant, and none of the point estimates suggest that a merger reduces mortality by more than 10%. We find that one merger resulted in

[46] We report marginal effects for this model in table C-9, alongside the marginal effects estimates from the fully specified model.

[47] We pool several of the smaller acquisitions from the sample into a single group.

FIGURE 1.—MARGINAL EFFECT ESTIMATES SEPARATELY BY ACQUISITION: FULL MATCHING SAMPLE MORTALITY



Marginal effect estimates are reported as percentages/100. The effects are combined for the following mergers (open circle): Butler/DiCuccio; Good Samaritan/NY Institute; Texas Children's/Women's Specialists'; Christ Hospital/Hyde Park Internists; Scripps Health/Penn Elm and Theda Care/Nelson Family.

statistically significant increased mortality, an effect that is greater than 10% but imprecisely estimated. Taken together, these results suggest that mergers may possibly improve mortality modestly but also have the possibility of substantially worsening it.

Next, we consider health outcomes that are more closely related to ambulatory treatment. Figure 2 considers acquisition-specific marginal effects of outcomes related to diabetes. Panels a and b correspond to marginal effect estimates of a provider acquisition on asymptomatic and symptomatic diabetes complications, respectively. Panel C considers the effect of acquisitions on glaucoma, an eye condition that may result from the progression of diabetes. We find that the vast majority of acquisitions have no statistically significant effect on these outcomes. One merger improves symptomatic diabetes conditions, but two mergers result in worse symptomatic diabetes conditions. None of the mergers have statistically significant effects on the other outcomes.

Next, we consider health outcomes associated with the treatment of hypertension, perhaps the most prevalent chronic condition in the sample. Figure 3 presents the effects of acquisitions on AMI in panel a, ischemic heart disease in panel b, and acute cardiac conditions, which includes AMI and stroke, in panel c. Again, we rarely observe that mergers have statistically significant effects on any of our health outcomes. In panel a, we observe that none of the acquisitions result in fewer heart attacks. In panel b, we observe that none of the acquisitions reduce the hazard for ischemic heart disease. In panel c, two mergers improved patient's progression into acute cardiac conditions, and one merger resulted in worse acute cardiac outcomes.

The effects from our hypertension samples are especially relevant since our sample of acquisitions includes many cardiology groups, which we identify in figure 3 using blocks as symbols rather than dots. Although integration should theoretically affect any type of physician who treats his or her observed patients, one might expect that the largest impact would be realized among patients with conditions treated by the specialty of the acquired physicians.[48] The effects of cardiology transactions appear to be found uniformly throughout the distribution of the estimated effects. The extent to which transactions do not improve hypertensive patient outcomes suggests that our findings are unlikely attributable to choosing irrelevant health conditions for the specialties involved in the acquisitions considered.

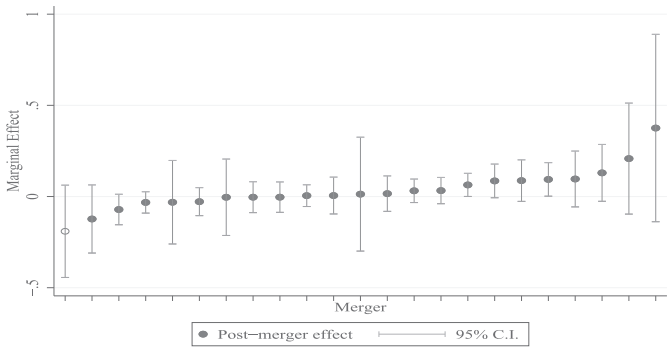
C. *Quarter-Specific Effects*

Although our overall and acquisition-specific results provide little evidence that full financial integration of hospitals and physicians leads to improved health outcomes, perhaps the development of best practices, training, and the acculturation of physicians takes time to accomplish. If so, the average effect for the period following an acquisition might mask potential benefits that do not become realized until months or years following the acquisition. We consider the potential for these types of period-specific effects by plotting quarter-specific acquisition marginal effects from before and after an acquisition. We also provide the respective 5% confidence intervals to identify statistically significant effects.

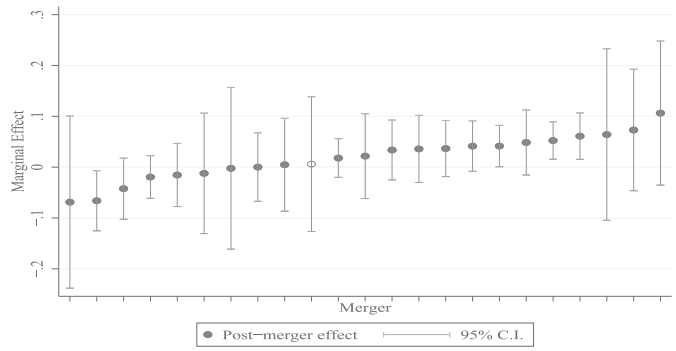
Figure 4 considers mortality estimates for each period in the full sample following a merger separately for men and women. If integration takes time to achieve clinical benefits,

[48]Conceptually, integration could allow providers to coordinate care and ensure better outcomes across a variety of treatments, regardless of whether the relevant specialist directly treats the condition.

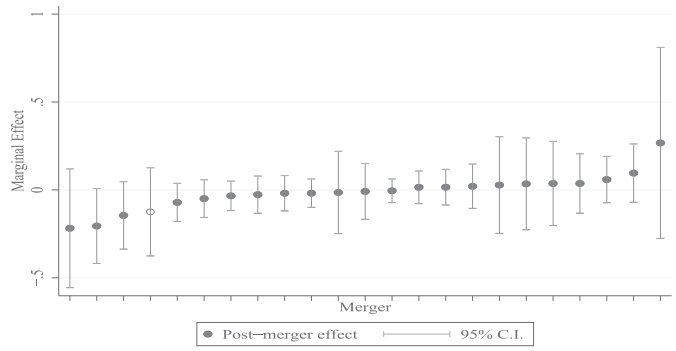
FIGURE 2.—MARGINAL EFFECTS SEPARATELY BY ACQUISITION: DIABETES



(a) Asymptomatic complications



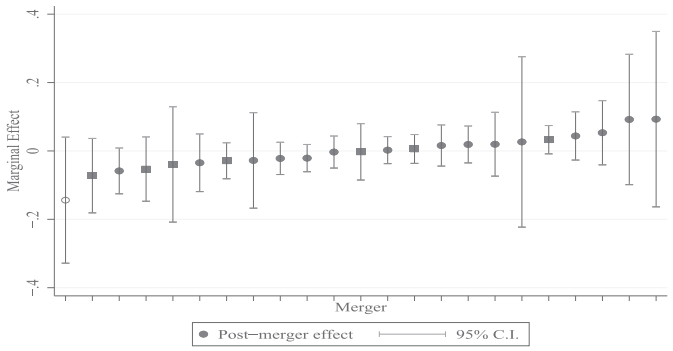
(b) Symptomatic complications



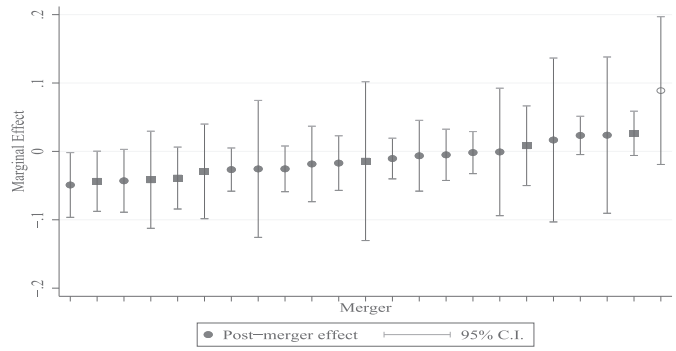
(c) Glaucoma

Marginal effect estimates are reported as percentages/100. The effects are combined for the following mergers (open circle): Butler/DiCuccio; Good Samaritan/NY Institute; Texas Children's/Women's Specialists; Christ Hospital/Hyde Park Internists; Scripps Health/Penn Elm, and Theda Care/Nelson Family.

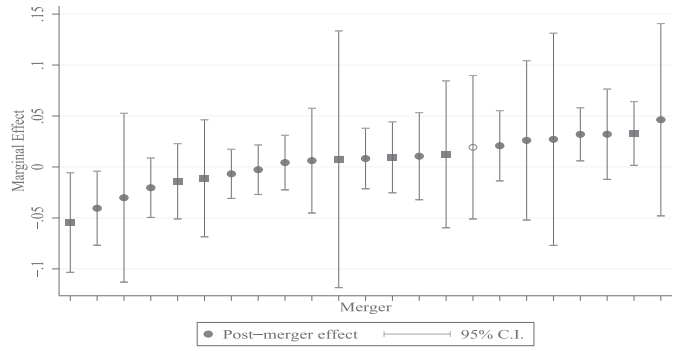
FIGURE 3.—MARGINAL EFFECTS SEPARATELY BY ACQUISITION: HYPERTENSION



(a) AMI



(b) Ischemic heart disease



(c) Acute cardiac conditions

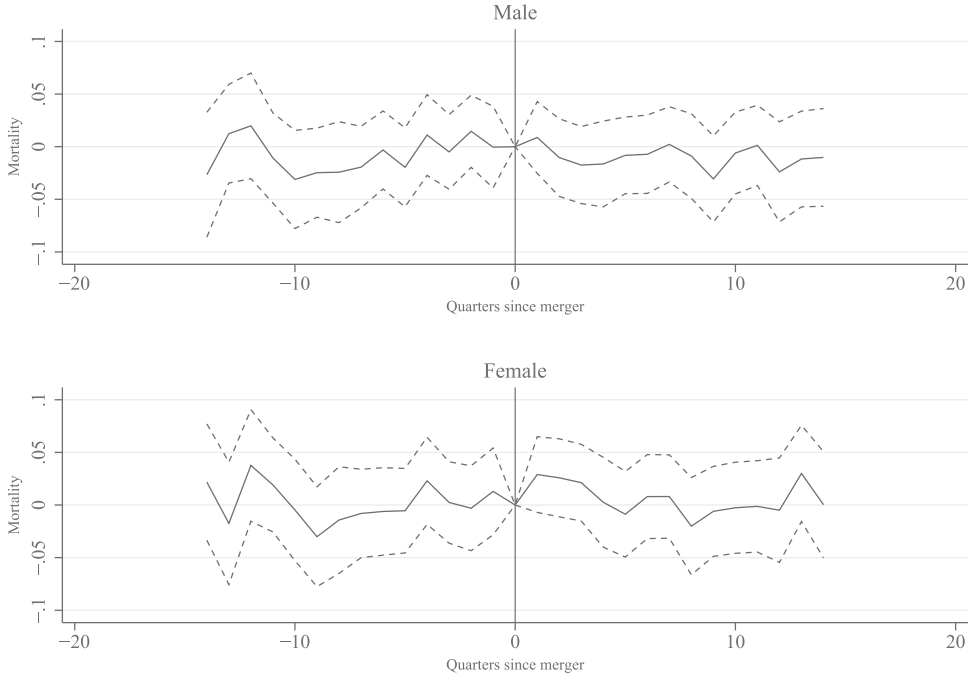
Marginal effect estimates are reported as percentages/100. The effects are combined for the following mergers (open circle): Butler/DiCuccio; Good Samaritan/NY Institute; Texas Children's/Women's Specialists; Christ Hospital/Hyde Park Internists; Scripps Health/Penn Elm, and Theda Care/Nelson Family. Cardiology transactions are square indicators.

we would expect that the quarter-specific acquisition coefficients would be increasing in magnitude (i.e., by becoming more negative) in the periods following the acquisition. We do not find any evidence of such a trend.

Next, we consider the period-specific acquisition effects for diabetes complications in figure 5, separately for men and women. Again, we would expect merger effects to increase over time (i.e., become more negative) if mergers took time to achieve clinical benefits with respect to these measures. However, the marginal effect estimates are generally flat for every diabetes-related outcome that we consider for both men and women.

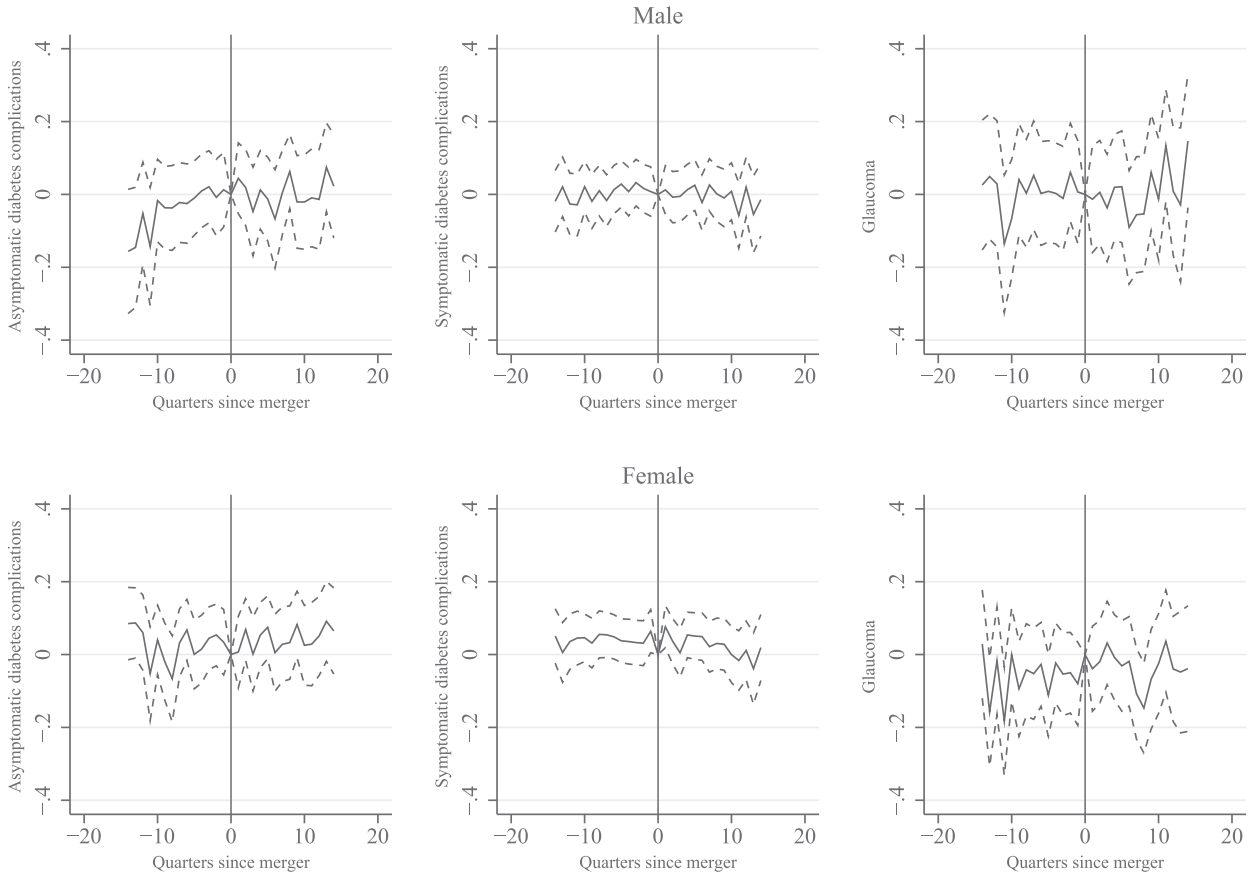
Finally, figure 6 considers the set of hypertension conditions. These results are the most precisely estimated, and perhaps the most relevant, given the importance of cardiologists

FIGURE 4.—PRE- AND POSTMERGER MARGINAL EFFECTS: MORTALITY



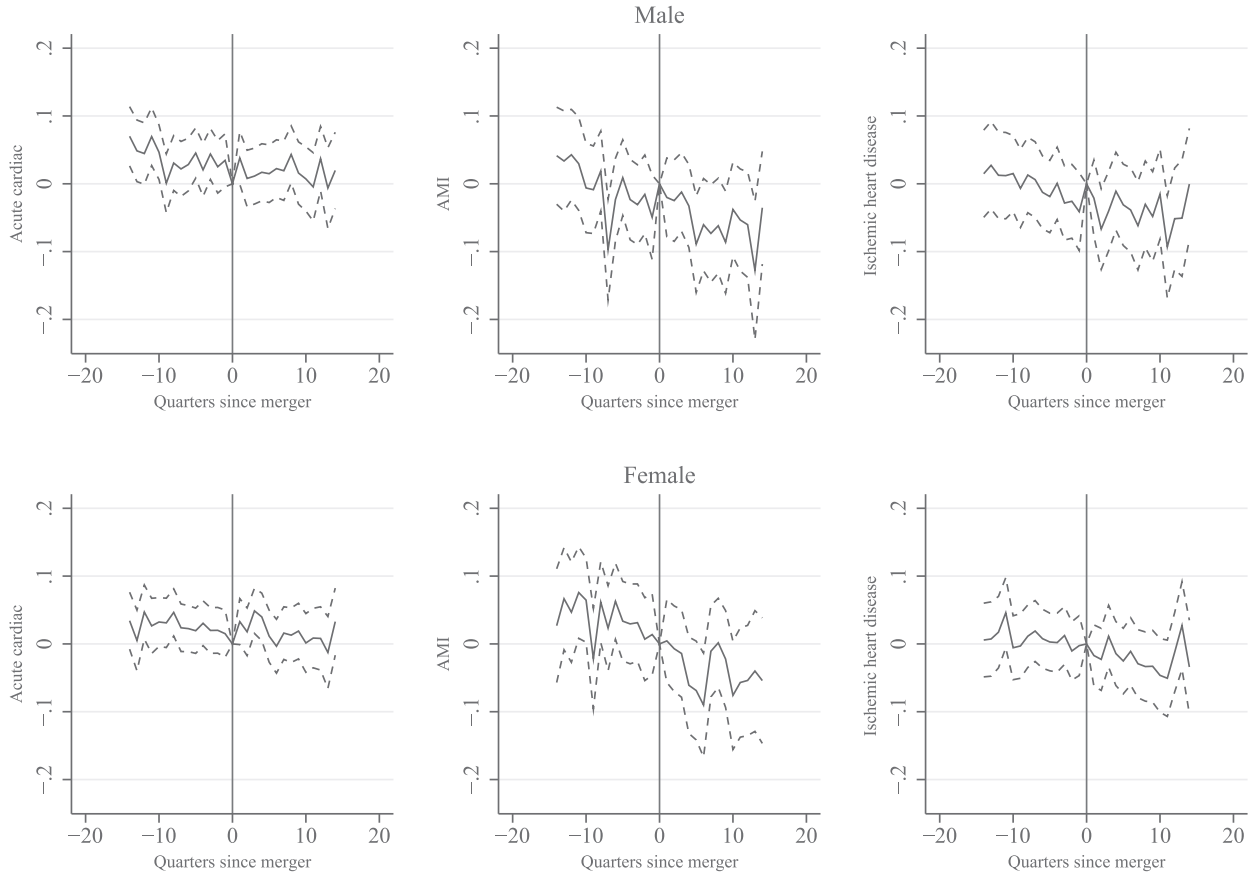
Marginal effect estimates are reported as percentages/100 on the *y*-axis.

FIGURE 5.—PRE- AND POSTMERGER MARGINAL EFFECTS: DIABETES



Marginal effect estimates are reported as percentages/100 on the *y*-axis.

FIGURE 6.—PRE- AND POSTMERGER MARGINAL EFFECTS: HYPERTENSION



Marginal effect estimates are reported as percentages/100 on the *y*-axis.

to our sample. Some of the estimates suggest a pattern of declining outcome probabilities relative to a control group, especially for AMI and ischemic heart disease. However, these patterns appear to have begun prior to the merger, consistent with targeted practice groups improving over time differently from the control group. Perhaps this difference is responsible for the limited effects we found for these outcomes in our full controls matching specification. However, the patterns here are largely statistically insignificant, suggesting little evidence that clinical benefits take time to become realized in the treatment of hypertension.

VI. Conclusion

We consider the health consequences of 28 physician acquisitions by hospitals using Medicare claims and a sample of acquisitions. We focus on a broad range of health outcomes, each with different analytical strengths and weaknesses. Some measures, such as AMI and mortality, are easily observed and are consistently identified across time and providers. Other measures that are not as easily diagnosed, such as our diabetes outcomes, have the advantage that they are directly related to the treatment of the disease. We find little evidence that physician integration into hospital systems affects the health outcomes (as we measure them) of Medicare beneficiaries. We interpret our results as implying that acquisitions of this type have small clinical benefits related to the treatment of hypertension and diabetes.

One possible limitation of our study is that some of the vertical acquisitions that we consider may also increase physician concentration. If so, prior research has demonstrated that increased horizontal concentration may lessen competition for quality (Koch et al., 2017) and thus offset efficiencies associated with vertical integration. Given that some of our mergers involve small physician practices or take place in a large metropolitan area, we do not believe that confounding horizontal effects are responsible for our overall results.

Another limitation is our focus on clinical efficiencies related to specific health conditions. It is possible that the transactions could lead to efficiencies for other health conditions or along other dimensions of performance (i.e., cost efficiencies) that we do not consider. For this reason, our results do not allow us to reach conclusions about total welfare. Despite these limitations, we emphasize that the health outcomes that we consider are important and prevalent. Nearly everyone in our Medicare sample becomes hypertensive, and over one-third of Medicare patients become diabetic. These conditions can have serious health consequences if not properly managed,

and they can interact with the treatment and development of other types of health conditions. This would appear to create the possibility of benefits from integration. Despite the opportunities for these potential benefits, we find little evidence they were realized in the transactions we consider.

REFERENCES

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies* 72:1 (2005), 1–19. https://doi.org/10.1111/0034-6527.00321

Afendulis, Christopher C., and Daniel P. Kessler, "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care," *American Economic Review* 97 (2007), 1013–1020. https://doi.org/10.1257/aer.97.3.1013, PubMed: 26236038

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler, "Vertical Integration: Hospital Ownership of Physician Practices Is Associated with Higher Prices and Spending," *Health Affairs* 33 (2014), 756–763. https://doi.org/10.1377/hlthaff.2013.1279, PubMed: 24799571

Baker, Laurence C., M. Kate Bundorf, Anne B. Royalty, and Zachary Levin, "Physician Practice Competition and Prices Paid by Private Insurers for Office Visits," *JAMA* 312 (2014), 1653–1662. 10.1001/jama.2014.10921

Blankenthorn, Dana, "Vertical Integration the Answer to Health Care Costs," SmartPlanet.com (2009). https://doi.org/10.1377/hlthaff.2013.1279

Bresnahan, Timothy F., and Jonathan D. Levin, "Vertical Integration and Market Structure," NBER technical report (2012). https://doi.org/10.3386/w17889

Burns, L. R., Jeff C. Goldsmith, and Aditi Sen, "Horizontal and Vertical Integration of Physicians: A Tale of Two Tails," *Advances in Health Care Management* 15 (2013), 39–117. https://doi.org/10.1108/s1474-8231(2013)0000015009, PubMed: 24749213

Burns, Lawton Robert, and Ralph W. Muller, "Hospital-Physician Collaboration: Landscape of Economic Integration and Impact on Clinical Integration," *Milbank Quarterly* 86 (2008), 375–434. https://doi.org/10.1111/j.1468-0009.2008.00527.x, PubMed: 18798884

Capps, Cory, David Dranove, and Christopher Ody, "The Effect of Hospital Acquisitions of Physician Practices on Prices and Spending," mimeograph (2015). https://doi.org/10.1016/j.jhealeco.2018.04.001, PubMed: 29727744

——— "Physician Practice Consolidation Driven by Small Acquisitions, So Antitrust Agencies Have Few Tools to Intervene," Health Affairs 36 (2017), 1556–1563. https://doi.org/10.1377/hlthaff.2017.0054, PubMed: 28874481

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47 (1980), 225–238. https://doi.org/10.2307/2297110

Ciliberto, Federico, and David Dranove, "The Effect of Physician–Hospital Appliations on Hospital Prices in California," *Journal of Health Economics* 15 (2006), 29–38. https://doi.org/10.1016/j.jhealeco.2005.04.008, PubMed: 16352360

Cooper, James C., Luke M. Froeb, Dan O'Brien, and Michael G. Vita, "Vertical Antitrust Policy as a Problem of Inference," *International Journal of Industrial Organization* 23 (2005), 639–664. https://doi.org/10.1016/j.ijindorg.2005.04.003

Cuellar, Alison Evans, and Paul J. Gertler, "Strategic Integration of Hospitals and Physicians," *Journal of Health Economics* 25 (2006), 1–28. https://doi.org/10.1016/j.jhealeco.2005.04.009, PubMed: 16309766

Dranove, David, and Christopher Ody, "Employed for Higher Pay: How Medicare Facility Fees Affect Hospital Employment of Physicians," unpublished paper (2016). https://doi.org/10.1257/pol.20170020

Forbes, S. J., and M. Lederman, "Does Vertical Integration Affect Firm Performance? Evidence from the Airline Industry," *RAND Journal of Economics* 41 (2010), 765–790. https://doi.org/10.1111/j.1756-2171.2010.00120.x

Gaynor, Martin, and William B. Vogt, "Antitrust and Competition in Health Care Markets" (pp. 1405–1487), in A. J. Cuyler and J. P. Newhouse, eds., *Handbook of Health Economics* (Amsterdam: Elsevier, 2000). https://doi.org/10.3386/w7112

Gottlieb, S., "The Doctor Won't See You Now: He's Clocked Out," *Wall Street Journal*, March 14, 2013.

Imbens, Guido W., and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences* (Cambridge: Cambridge University Press, 2015). https://doi.org/10.1017/CBO9781139025751

Imbens, Guido W., and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature* 47 (2009), 5–86. https://doi.org/10.1257/jel.47.1.5

Jin, G. Z., and P. Leslie, "Reputational Incentives for Restaurant Hygiene," *American Economic Journal: Microeconomics* 1 (2009), 237–267. https://doi.org/10.1257/mic.1.1.237

Kane, C. K., and D. W. Emmons, "New Data on Physician Practice Arrangements: Private Practice Remains Strong Despite Shifts toward Hospital Employment," AMA Policy Research Perspectives technical report (2013).

Keating, Nancy L., Mary Beth Landrum, Bruce E. Landon, John Z. Ayanian, Catherine Borbas, Robert Wolf, and Edward Guadagnoli, "The Influence of Physicians' Practice Management Strategies and Financial Arrangements on Quality of Care among Patients with Diabetes," *Medical Care* 42 (2004), 829–839. https://doi.org/10.1097/01.mlr.0000135829.73795.a7, PubMed: 15319608

Koch, Thomas G., Brett W. Wendling, and Nathan E. Wilson, "How Vertical Integration Affects the Quantity and Cost of Care for Medicare Beneficiaries," *Journal of Health Economics* 52 (2017), 19–32. https://doi.org/10.1016/j.jhealeco.2016.12.007, PubMed: 28182998

Kosova, R., F. Lafontaine, and R. Perrigot, "Organizational Form and Performance: Evidence from the Hotel Industry," this REVIEW 95 (2013), 1303–1323. https://doi.org/10.2139/ssrn.1034082

McWilliams, J. M., "Changes in Health Care Spending and Quality for Medicare Beneficiaries Associated with a Commercial ACO Contract," *JAMA* 310 (2013), 829–836. https://doi.org/10.1001/jama.2013.276302, PubMed: 23982369

Medicare Payment Advisory Commission, "Report to the Congress: Medicare Payment Policy," technical report (March 2012).

Mullainathan, S., and D. Scharfstein, "Do Firm Boundaries Matter?" *American Economic Review* 91 (2001), 195–199. https://doi.org/10.1016/j.jfineco.2013.11.001

Neprash, Hannah T., Michael E. Chernew, Andrew L. Hicks, Teresa Gibson, and J. Michael McWilliams, "Association of Financial Integration between Physicians and Hospitals with Commercial Health Care Prices," *JAMA Internal Medicine* 175 (2015), 1932–1939. https://doi.org/10.1001/jamainternmed.2015.4610

Novak, S., and S. Stern, "How Does Outsourcing Affect Performance Dynamics? Evidence from the Automobile Industry," *Management Science* 54 (2008), 1963–1979. https://doi.org/10.1287/mnsc.1080.0922

Post, Brady, Tom Buchmueller, and Andrew M. Ryan, "Vertical Integration of Hospitals and Physicians: Economic Theory and Empirical Evidence on Spending and Quality," *Medical Care Research and Review* 75 (2017), 399–433. https://doi.org/10.1177/1077558717727834, PubMed: 29148355

Rebitzer, James B., and Mark Votruba, "Organizational Economics and Physician Practices," NBER working paper 17535 (2011). http://www.nber.org/papers/w17535

Rosenbaum, Paul R., and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70 (1983), 41–55. https://doi.org/10.1093/biomet/70.1.41

——— "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," American Statistician 39 (1985), 33–38. https://doi.org/10.1080/00031305.1985.10479383

Vita, M. G., "Regulatory Restrictions on Vertical Integration and Control: The Competitive Impact of Gasoline Divorcement Policies," *Journal of Regulatory Economics* 18 (2000), 217–233. https://doi.org/10.1023/A:1008150819999

Whinston, Michael D., *Lectures on Antitrust Economics* (Cambridge, MA: MIT Press, 2006).

Wilson, Nathan E., "Vertical Separation Increases Gas Prices," *Economic Inquiry* 53 (2015).

Wolinsky, Asher, "Competition in a Market for Informed Experts' Services," *RAND Journal of Economics* (1993), 380–398. https://doi.org/10.2307/2555964