Impact of Hospital Affiliation With a Flagship Hospital System on Surgical Outcomes

Omar I. Ramadan, MD, MSc,†✉ Paul R. Rosenbaum, PhD,†‡ Joseph G. Reiter, MS,§ Siddharth Jain, DrPH,†§ Alexander S. Hill, BS,§ Sean Hashemi, BS,§ Rachel R. Kelz, MD, MSCE, MBA,*† Lee A. Fleisher, MD,†‖¶ and Jeffrey H. Silber, MD, PhD†§#***

Objective: To compare general surgery outcomes at flagship systems, flagship hospitals, and flagship hospital affiliates versus matched controls.

Summary Background Data: It is unknown whether flagship hospitals perform better than flagship hospital affiliates for surgical patients.

Methods: Using Medicare claims for 2018 to 2019, we matched patients undergoing inpatient general surgery in flagship system hospitals to controls who underwent the same procedure at hospitals outside the system but within the same region. We defined a "flagship hospital" within each region as the major teaching hospital with the highest patient volume that is also part of a hospital system; its system was labeled a "flagship system." We performed 4 main comparisons: patients treated at any flagship system hospital versus hospitals outside the flagship system; flagship hospitals versus hospitals outside the flagship system; flagship hospital affiliates versus hospitals outside the flagship system; and flagship hospitals versus affiliate hospitals. Our primary outcome was 30-day mortality.

Results: We formed 32,228 closely matched pairs across 35 regions. Patients at flagship system hospitals (32,228 pairs) had lower 30-day mortality than matched control patients [3.79% vs. 4.36%, difference = −0.57% (−0.86%, −0.28%), $P < 0.001$]. Similarly, patients at flagship hospitals (15,571/32,228 pairs) had lower mortality than control patients. However, patients at flagship hospital affiliates (16,657/32,228 pairs) had similar mortality to matched controls. Flagship hospitals had lower mortality than affiliate hospitals [difference-in-differences = −1.05% (−1.62%, −0.47%), $P < 0.001$].

Conclusions: Patients treated at flagship hospitals had significantly lower mortality rates than those treated at flagship hospital affiliates. Hence, flagship system affiliation does not alone imply better surgical outcomes.

Key words: flagship hospital, general surgery, health policy, hospital brand, hospital quality, hospital system, Medicare, multimorbidity, surgical outcomes

(*Ann Surg* 2024;279:631–639)

Although primarily driven by financial and regulatory considerations,[1,2] hospital system mergers and acquisitions are typically presented to patients and policymakers as beneficial to patient care,[3,4] such as through improved care coordination across hospitals[5,6] or economies of scale that facilitate process efficiency and investments in advanced technology.[5–7] Perhaps most fundamentally, affiliation with a major regional hospital system is often promoted by these same systems as an opportunity for patients to benefit locally from the same standard of care experienced at their flagship hospital.[8,9] This branding association with one of the best hospitals in the region offers affiliates the ability to distinguish themselves from surrounding hospitals.

Indeed, there is widespread patient perception,[8] perhaps misguided,[10–12] that surgical care provided at hospitals affiliated with a major regional system is no different than care at the system's flagship hospital, suggesting better surgical care at affiliated hospitals than hospitals outside the flagship system. Prior research has focused on what affiliation offers for hospitals by looking at outcomes and financial performance before and after affiliation.[5,13–15] However, a vital question remains: should patients and policymakers expect superior surgical outcomes from hospitals affiliated with major regional systems (ie, "flagship systems"), particularly hospitals other than the flagship hospital, compared with hospitals outside the flagship system?

To explore this, we performed a matched cohort study in 35 of the nation's largest hospital referral regions (HRRs)[16,17] using Medicare claims data to compare surgical outcomes between patients at: (1) flagship system hospitals (ie, all hospitals within the preeminent regional academic hospital system) versus hospitals within the same HRR but outside the flagship system; (2) flagship hospitals versus within-HRR hospitals outside the flagship system; and (3) affiliate hospitals in flagship systems versus within-HRR hospitals outside the flagship system.

From the *Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; †Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA; ‡Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA; §Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA; ‖Center for Perioperative Outcomes Research and Transformation, University of Pennsylvania, Philadelphia, PA; ¶Department of Anesthesiology and Critical Care, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; #Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; and **Department of Health Care Management, The Wharton School, University of Pennsylvania, Philadelphia, PA.

✉omar.ramadan@pennmedicine.upenn.edu.

R.P.R., F.L.A., and S.J.H.: conception or design. R.O., R.P.R., R.J.G., J.S., H.A.S., H.S., K.R.R., F.L.A., and S.J.H.: acquisition, analysis, or interpretation of data. R.O., R.P.R., F.L.A., and S.J.H.: drafting of the manuscript. R.O., R.P.R., R.J.G., J.S., H.A.S., H.S., K.R.R., F.L.A., and S.J.H.: critical revision of the manuscript for important intellectual content and final approval of the version to be published. R.P.R., R.J.G., and S.J.H.: statistical analysis. F.L.A. and S.J.H.: obtaining funding. J.S. and S.J.H.: administrative, technical, or material support. S.J.H.: supervision.

This research was funded by a grant from the National Institute on Aging (Grant # R01AG060928). Dr. Ramadan was supported by the Agency for Healthcare Research and Quality National Research Service Award (T32; Grant # 5T32HS026116).

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.annalsofsurgery.com.

ISSN: 0003-4932/24/27904-0631
DOI: 10.1097/SLA.0000000000006132

METHODS

Patient Population

We used Medicare administrative claims data (Inpatient, Outpatient, Carrier/Part B, Skilled Nursing Facility, Home Health Agency, and Durable Medical Equipment files) for all fee-for-service Medicare beneficiaries through the Centers for Medicare and Medicaid Services (CMS) Virtual Research Data Center.[18] We analyzed patients 66 years of age and older who underwent inpatient general surgery procedures between 2018 and 2019. These patients were categorized into clinically relevant groups by the International Classification of Diseases, Tenth Revision (ICD-10) principal procedure codes (Supplemental Digital Content 1, Section A eTable 1, http://links.lww.com/SLA/E923). For patients with multiple procedures, we used their first. We excluded patients if, in a 1-year lookback before their admission, they either: (1) lacked fee-for-service Medicare claims; (2) did not have complete enrollment in Medicare Parts A and B; or (3) were enrolled in a health maintenance organization at any point.

Defining Hospital Systems

The Agency for Healthcare Research and Quality maintains a database of health systems across the United States, which it defines as including "at least one hospital and at least one group of physicians that provid[e] comprehensive care…who are connected with each other and with the hospital through common ownership or joint management."[16] Using this list for 2018, we defined "hospital systems" as health systems with at least 2 acute care hospitals within the same HRR. HRRs are geographic areas that share the same tertiary care referral patterns based on Medicare data, reflecting distinct health care markets.[17] HRRs have been widely used to describe health care utilization and cost.[14,19,20] Of 306 HRRs in the United States, 35 met our volume ($N \geq 20{,}000$) and system criteria, as summarized below.

We defined a "major Council of Teaching Hospitals and Health Systems (COTH) hospital" as a hospital with a resident-to-bed ratio ≥ 0.25—consistent with "major" or "very major" teaching hospitals—that is also a member of the COTH. Within each HRR, we defined the "flagship hospital" as the largest (ie, highest combined medical and surgical patient volume) major COTH hospital that also had affiliated hospitals within the same HRR. We defined the "flagship system" as the system that included the flagship hospital. Thus, each HRR was defined to have only 1 flagship hospital and 1 flagship system. Patients at all other hospitals within the same HRR but not in the flagship system were labeled as potential controls. Potential controls could therefore come from unaffiliated hospitals, hospitals affiliated with other academic centers within the same HRR, or even other major COTH hospitals within the same HRR but not in the flagship system. We investigate this further in the stability analysis described below. See Supplemental Digital Content 1, Section B and eTables 2, http://links.lww.com/SLA/E923 and Section B and eTables 3, http://links.lww.com/SLA/E923 for further detail.

HRRs With Multiple Major COTH Hospitals

Some HRRs contained several major COTH hospitals. In our primary analysis, patients from all such major COTH hospitals that were not labeled part of the flagship system per our criteria were instead included as potential controls. However, we also performed a stability analysis that removed matched pairs containing control patients admitted to these nonflagship system major COTH hospitals. As will be seen, this stability analysis only strengthened our findings.

Defining Multimorbidity

In our previous work,[21] we defined multimorbidity for older surgical patients as the presence of at least 1 cluster of comorbidities—termed qualifying comorbidity sets—confidently associated with at least double the odds of 30-day mortality compared with the typical patient undergoing the same procedure in the same age group. We have since refined our multimorbidity definition[22] to be compatible with ICD-10 codes and incorporate functional status indicators, allowing us to identify particularly high-risk patients. These updates were completed before the present study and utilized data that did not overlap with this study.

Outcomes

The primary outcome was 30-day mortality. We also examined 90-day mortality and an updated 30-day failure-to-rescue outcome, which represents mortality after in-hospital postoperative complications.[23–26] An updated list of complications used for computing failure-to-rescue is provided in Supplemental Digital Content 1, Section C eTable 4, http://links.lww.com/SLA/E923.

Statistical Analysis

Matching Methodology

We used optimal subset matching[27–30] to balance many covariates in an optimal manner[31] to match patients in flagship system hospitals ("treated" hospitals) to control patients in other hospitals outside the flagship system who underwent the same procedure within the same HRR. This required an exact match for the surgical procedure and, subject to that requirement, picked the closest possible pairing of patients based on patient demographics, socioeconomic status (including dual-eligibility and neighborhood education and poverty levels), presence of a multimorbid qualifying comorbidity set, emergent admission status, and risk of death, for a total of 147 risk factors as displayed in Supplemental Digital Content 1, Section D eTable 5, http://links.lww.com/SLA/E923. Matching was performed at the patient level. To further strengthen our match quality, we aimed to attain standardized differences <0.1 after matching, more stringent than the conventional <0.2.[32,33] Matching was completed before viewing the outcomes.[34]

Comparing Hospitals

We performed 4 main analyses using the one matched sample we described above, which paired patients in flagship hospital systems to control patients undergoing the same inpatient surgical procedure within the same HRR but at a hospital outside the flagship system. Of note, all of our analyses look at these pairs, sometimes grouping pairs in different ways depending on the question, although who is paired with whom never changes. First, we examined all matched pairs to ask whether patients have lower mortality at any flagship system hospitals than at a hospital outside the flagship system. Then, we separated this pool of flagship system hospital-matched pairs into flagship hospital-matched pairs and affiliate hospital-matched pairs. This allowed us to address 3 additional questions, referred to as analyses 2, 3, and 4: (2) do patients have lower mortality at a flagship hospital than a hospital outside the flagship system; (3) do patients have lower mortality at an affiliate of the flagship hospital (excluding the flagship hospital itself) than a hospital outside the flagship system; and (4) do patients have lower mortality at flagship hospitals versus affiliate hospitals?

TABLE 1. Quality of Matched Pairs Comparing Patients, Before and After Matching, in Flagship Hospital Systems to Their Matched Controls* From Other Hospitals in the Same HRR

	Flagship systems		Matched controls		Standardized difference	
Variable	**Before**	**After**	**After**	**Before**	**Before**	**After**
N	37,223	32,228	32,228	97,728	–	–
Demographics (%)						
Age on day of surgery (mean years)	75.22	75.41	75.63	75.79	−0.08	−0.03
Age 85+ (%)	12.2	13.0	13.0	14.0	−0.05	0.00
Race						
White non-Hispanic	84.0	84.2	85.3	85.3	−0.04	−0.03
Black	9.5	9.2	8.7	7.8	0.06	0.02
Hispanic	0.9	1.0	0.6	1.3	−0.04	0.04
Female	55.8	56.1	56.7	57.1	−0.02	−0.01
Dually eligible	12.1	12.3	12.5	14.4	−0.07	−0.01
High poverty neighborhood†	7.8	7.9	7.0	8.4	−0.02	0.03
Low education neighborhood‡	10.4	10.4	9.8	13.2	−0.09	0.02
Emergent admission	39.7	42.3	43.9	46.6	−0.14	−0.03
Probability of death on admission (%)	5.0	5.0	5.0	4.9	0.01	0.01
Comorbidities (%) (see Supplement* for full list)						
Chronic pulmonary diseases	27.2	26.8	27.6	27.3	0.00	−0.02
Diabetes with complications	26.0	25.2	25.7	26.2	−0.01	−0.01
Heart failure	25.0	25.0	25.4	25.8	−0.02	−0.01
Protein calorie malnutrition	18.7	18.6	17.0	15.2	0.09	0.04
Thrombocytopenia and other hematological disorders	16.1	16.0	14.4	14.7	0.04	0.04
Metastatic cancers	15.7	15.0	13.9	12.7	0.09	0.03
CKD stage 4–5 or dialysis	6.1	6.1	5.7	5.9	0.01	0.02
Functional status (%)						
Home oxygen use	3.9	3.8	4.5	4.4	−0.02	−0.03
Home hospital bed or wheelchair use	2.1	1.9	2.2	2.4	−0.02	−0.01
Multimorbid (%)	60.3	60.5	60.6	59.3	0.02	0.00

*The 32 general surgery procedure groups were exactly matched. For a complete list of all 147 matching variables in all matches, see Supplemental Digital Content 1, Section D eTables 5–7.

†Proportion of patients that live in a zip code in which >20% of adults live below the federal poverty line.

‡Proportion of patients that live in a zip code in which <80% of adults have a high school diploma.

CKD indicates chronic kidney disease; diff ave, average difference.

Finally, based on literature suggesting higher-quality hospitals have superior outcomes for higher-risk surgical patients,[21,24,35–40] we separated pairs of patients with multimorbidity[22] from pairs without multimorbidity and asked whether those with multimorbidity derived a disproportionate mortality benefit in flagship systems compared with those without multimorbidity within each hospital type comparison.

Most of our analyses refer to a difference in 2 binary survival rates in matched pairs; see Fleiss et al.[41] A difference-in-difference is the difference of 2 such differences, and its variance is the sum of their 2 variances, by independence. Tables 3 and 4 report separate multiplicity adjustments to *P*-values using the Bonferroni-Holm method.[42,43]

The difference between 2 failure-to-rescue rates is considerably more complicated. A failure is a death following a complication. One person in a pair may have a complication when the other does not. As a consequence, the failure rate describes the population of pairs and is not meaningful for a single pair. We therefore created a standard error of the difference in failure rates by bootstrapping (ie, resampling) whole pairs.[44]

All analyses were completed using SAS version 9.4.[45] The study was approved by the Children's Hospital of Philadelphia Institutional Review Board.

RESULTS

Beginning with 37,223 general surgery patients in flagship system hospitals and 97,728 control patients in hospitals not in the flagship system but in the same HRR, we formed 32,228 closely matched pairs across 35 of the largest HRRs. As seen in Table 1 and Supplemental Digital Content 1, eTables 5–8, http://links.lww.com/SLA/E923, we achieved excellent matches. All 32 general surgical procedures were matched exactly. All demographic, socioeconomic status, comorbidity, and risk of death variables had absolute standardized differences below 0.1, usually considerably smaller.

Compared with their matched controls, flagship system hospitals were nearly twice as large, were more likely to be teaching hospitals, provide more advanced interventions, and have superior nursing resources (Table 2). However, these differences were largely attributable to the flagship hospitals themselves, whereas affiliate hospitals were far more similar to the matched controls. For instance, the mean number of beds at flagship hospitals was 974 versus 405 for matched controls, compared with 407 at affiliate hospitals versus 350 for matched controls.

Flagship System Hospitals Versus Hospitals Outside the Flagship System

Patients at flagship system hospitals had significantly lower rates of 30-day mortality than matched controls, who had highly similar comorbidities and socioeconomic status and underwent the same procedure within the same HRR [3.79% (flagship system hospitals) vs. 4.36% (controls), difference = −0.57% (95% CI: −0.86%, −0.28%), $P < 0.001$] (Table 3, Fig. 1a). Findings were similar for 90-day mortality [6.62% vs. 7.41%, difference = −0.79% (95% CI: −1.15%, −0.42%),

TABLE 2. Characteristics* of Hospitals in Flagship Systems, Flagship Hospitals, and Affiliate Hospitals in Flagship Systems Compared With Their Within-HRR-Matched Controls

	All US hospitals (N = 3407)	Flagship systems (N = 35)	Controls	Flagship hospitals (N = 35)	Controls	Affiliate hospitals (N = 121)	Controls
Study patients (N)	1,687,511	32,228	32,228	15,571	15,571	16,657	16,657
Number of beds (mean)	200.7	681	377	974	405	407	350
Teaching status							
Resident-to-bed ratio	0.09	0.369	0.215	0.593	0.254	0.160	0.179
COTH status (%)	7.2	57.2	24.9	100.0	30.2	17.2	20.0
Hospital resources (%)							
High technology status†	39.8	79.2	69.7	99.8	72.0	60.0	67.5
Availability of PCI‡	47.9	88.7	86.0	100.0	87.6	78.2	84.5
Comprehensive cardiac technology§	34.6	57.5	60.4	71.2	62.4	44.6	58.6
Nurse-to-bed ratio	1.63	1.78	1.72	2.06	1.79	1.53	1.66
Highest 1/3 (%)	33.3	49.3	47.5	68.0	50.1	31.8	45.1
Middle 1/3 (%)	33.3	34.7	33.7	22.9	32.9	45.7	34.4
Lowest 1/3 (%)	33.3	16.1	18.8	9.1	17.0	22.6	20.5
Nursing skill mix‖	0.89	0.965	0.937	0.978	0.939	0.953	0.934
Highest 1/3 (%)	33.3	62.9	48.8	73.8	51.1	52.7	46.6
Middle 1/3 (%)	33.3	30.8	36.9	23.3	35.2	37.8	38.5
Lowest 1/3 (%)	33.3	6.4	14.3	2.9	13.8	9.6	14.9

*Characteristics were weighted by the number of patients in each type of hospital.
†The proportion of patients in hospitals that perform both open heart surgery and organ transplantation.
‡The proportion of patients in hospitals that performed at least 10 PCIs during each year of the study.
§The proportion of patients in hospitals that have a cardiac catheterization laboratory, a coronary care unit, and provide cardiothoracic surgery services.
‖The proportion of registered nurses to the total number of registered nurses and licensed practical nurses.
COTH indicates Council of Teaching Hospitals and Health Systems; PCI, percutaneous coronary intervention.

$P < 0.001$] (Supplemental Digital Content 1, Section E eTable 9, http://links.lww.com/SLA/E923).

To determine whether care at flagship system hospitals was associated with a differential benefit for general surgery patients with-versus-without multimorbidity, we compared outcomes for these patients at flagship system hospitals versus other hospitals. Thirty-day mortality was lower for patients with multimorbidity treated at flagship system hospitals versus matched controls, while no significant difference was observed for patients without multimorbidity (Table 3, Fig. 2a). A larger mortality reduction was demonstrated for patients with-versus-without multimorbidity at flagship system hospitals versus matched controls [difference-in-differences = −0.81% (95% CI: −1.31%, −0.30%), $P = 0.002$]. Similar findings were seen for 90-day mortality (Supplemental Digital Content 1, eTable 9, http://links.lww.com/SLA/E923).

No differences were found in rates of in-hospital postoperative complications (Supplemental Digital Content 1, eTable 10, http://links.lww.com/SLA/E923). However, rates of 30-day failure-to-rescue were significantly lower in flagship system hospitals compared with matched controls [11.12% vs. 12.93%, difference = −1.81% (95% CI: −2.65%, −0.96%), $P < 0.001$] (Table 4). A larger reduction in 30-day failure-to-rescue was observed for patients with-versus-without multimorbidity at flagship system hospitals versus matched controls [difference-in-differences = −1.48% (95% CI: −2.96%, 0.00%), $P < 0.05$]; however, this P-value exceeds 0.05 after multiplicity adjustment.

Flagship Hospitals Versus Hospitals Outside the Flagship System

General surgery patients at flagship hospitals had lower rates of 30-day and 90-day mortality than matched controls outside the flagship system undergoing the same procedure within the same HRR [30-day: 3.30% (flagship hospitals) vs. 4.41% (controls), difference = −1.11% (95% CI: −1.53%, −0.70%), $P < 0.001$; 90-day: 6.17% vs. 7.62%, difference = −1.45% (95% CI: −1.98%, −0.92%), $P < 0.001$] (Table 3, Fig. 1b; Supplemental Digital Content 1, eTable 9, http://links.lww.com/SLA/E923). Patients with multimorbidity treated at flagship hospitals had lower 30-day mortality rates than matched controls [4.93% vs. 6.70%, difference = −1.77% (95% CI: −2.42%, −1.12%), $P < 0.001$], while no difference was observed for patients without multimorbidity (Fig. 2b). A larger mortality reduction was noted for patients with-versus-without multimorbidity at flagship hospitals versus matched controls [difference-in-differences = −1.67% (95% CI: −2.38%, −0.96%), $P < 0.001$]. The findings were similar for 90-day mortality.

Rates of 30-day failure-to-rescue were lower at flagship hospitals relative to matched controls [9.28% vs. 12.75%, difference = −3.47% (95% CI: −4.63%, −2.32%), $P < 0.001$] (Table 4). Again, this improvement was concentrated in patients with multimorbidity at flagship hospitals. Rates of in-hospital postoperative complications were similar but statistically higher in flagship hospitals versus matched controls [35.12% vs. 33.83%, difference = 1.28% (95% CI: 0.30%, 2.27%), $P = 0.01$], although no disproportionate difference was seen for patients with-versus-without multimorbidity (Supplemental Digital Content 1, eTable 10, http://links.lww.com/SLA/E923).

Flagship Hospital Affiliates Versus Hospitals Outside the Flagship System

Unlike for flagship systems as a whole or flagship hospitals specifically, general surgery patients at affiliate hospitals in flagship systems did not have significantly different rates of 30-day or 90-day mortality compared with control patients receiving the same procedure at within-HRR hospitals outside the flagship system [30-day: 4.25% (affiliate hospitals) vs. 4.32%

TABLE 3. Rates of 30-Day Mortality for General Surgery Patients in all Flagship System Hospitals, Flagship Hospitals Only, or Affiliated Hospitals in Flagship Systems Compared With Within-HRR-Matched Controls at Other Hospitals

	N	Flagship system hospitals (%)	Matched controls (%)	Difference in rates (%)	95% CI (%)	*P*	Adjusted *P*-value*
Flagship system hospitals versus matched controls							
All patients	32,228	3.79	4.36	–0.57	(—0.86, –0.28)	< 0.001	0.001
With multimorbidity	19,317	5.79	6.69	–0.90	(–1.36, –0.44)	< 0.001	0.001
Without multimorbidity	12,511	0.74	0.84	–0.10	(–0.32, 0.13)	0.43	1.000
With-versus-without multimorbidity (difference-in-differences)				–0.81	(–1.31, –0.30)	0.002	0.015
Flagship hospitals versus matched controls							
All patients	15,571	3.30	4.41	–1.11	(–1.53, –0.70)	< 0.001	< 0.001
With multimorbidity	9488	4.93	6.70	–1.77	(–2.42, –1.12)	< 0.001	< 0.001
Without multimorbidity	5852	0.70	0.80	–0.10	(–0.43, 0.23)	0.59	1.000
With-versus-without multimorbidity (difference-in-differences)				–1.67	(–2.38, –0.96)	< 0.001	< 1.000
		Affiliate hospitals (%)	Matched controls (%)	Difference in rates (%)	95% CI (%)	*P*	Adjusted *P*-value*
Affiliate hospitals in flagship systems versus matched controls							
All patients	16,657	4.25	4.32	–0.07	(–0.48, 0.35)	0.77	1.000
With multimorbidity	9829	6.62	6.68	–0.06	(–0.73, 0.61)	0.88	1.000
Without multimorbidity	6659	0.78	0.87	–0.09	(–0.41, 0.23)	0.63	1.000
With-versus-without multimorbidity (difference-in-differences)				0.03	(–0.70, 0.75)	0.94	1.000
		Flagship difference (%)	Control difference (%)	Difference-in-differences (%)	95% CI (%)	*P*	Adjusted *P*-value*
Difference-in-differences: flagship hospitals minus affiliate hospitals in flagship systems							
All patients		–0.95	0.10	–1.05	(–1.62, –0.47)	< 0.001	0.003
With multimorbidity		–1.69	0.02	–1.71	(–2.63, –0.79)	< 0.001	0.003
Without multimorbidity		–0.08	–0.07	–0.01	(–0.45, 0.43)	0.96	1.000

*Adjusted *P*-values use the Bonferroni-Holm method.[42,43]

(controls), difference = −0.07% (95% CI: −0.48%, 0.35%), $P = 0.77$; 90-day: 7.04% vs. 7.21%, difference = −0.17% (95% CI: −0.68%, 0.34%), $P = 0.53$] (Table 3, Fig. 1c; Supplemental Digital Content 1, eTable 9, http://links.lww.com/SLA/E923). Also, unlike flagship systems or flagship hospitals, no differences were observed for patients with or without multimorbidity at affiliate hospitals in flagship systems versus matched controls (Fig. 2c). Similar findings were observed for 30-day failure-to-rescue and in-hospital postoperative complications (Table 4; Supplemental Digital Content 1, eTable 10, http://links.lww.com/SLA/E923).

Flagship Hospitals Versus Affiliate Hospitals

To further compare surgical outcomes at flagship hospitals versus affiliate hospitals, we performed a difference-in-differences analysis comparing the performance of flagship hospitals versus their controls to that of affiliate hospitals versus their controls. This allowed us to compare flagship hospitals with affiliates. Rates of 30-day mortality and 30-day failure-to-rescue were significantly lower at flagship hospitals versus controls compared with affiliate hospitals versus controls [30-day mortality difference-in-differences = −1.05% (95% CI: −1.62%, −0.47%), $P < 0.001$; 30-day failure-to-rescue difference-in-differences = −3.37% (95% CI: −5.08%, −1.67%), $P < 0.001$] (Tables 3, 4). These findings were seen in patients with multimorbidity but not in those without multimorbidity. No significant difference-in-differences were observed for in-hospital complications (Supplemental Digital Content 1, eTable 10, http://links.lww.com/SLA/E923).

HRRs With Multiple Major COTH Hospitals: Stability Analysis

As aforementioned, some HRRs contained several major COTH hospitals (eg, Boston, MA) while others had only one (eg, Charlotte, NC). In an HRR with several COTH hospitals in different systems, a matched control patient may have come from a COTH hospital. Including major COTH controls could blunt the primary findings by comparing the outcomes of patients within the flagship system to those of control patients treated at major COTH hospitals. We performed a stability analysis that removed matched pairs containing control patients admitted to these nonflagship system major COTH hospitals, thereby examining the subset of matched pairs in which control patients were not from a major COTH hospital outside the flagship system. The stability analysis removed 18.7% of pairs (6024/32,228 matched pairs), with no pairs removed in 14 of the 35 HRRs in the study. In the stability analysis balance table (eTable 8), as expected, the resident-to-bed ratio from all controls after excluding COTH controls declined slightly (to 0.101 vs. 0.173 before exclusions).

In the stability analysis, our main findings were unchanged: we still found that flagship hospital patient outcomes were better than stability controls, patients at affiliate hospitals in the flagship system fared no better than controls, and the flagship hospital mortality benefit was entirely due to improved mortality for patients with multimorbidity (Supplemental Digital Content 1, eTables 8, 11, http://links.lww.com/SLA/E923). Our conclusions were therefore very stable, and this appears to reflect the fact that 14 HRRs had only one qualifying COTH system, and even when there was a second COTH system, most patients in the HRR were not treated there.

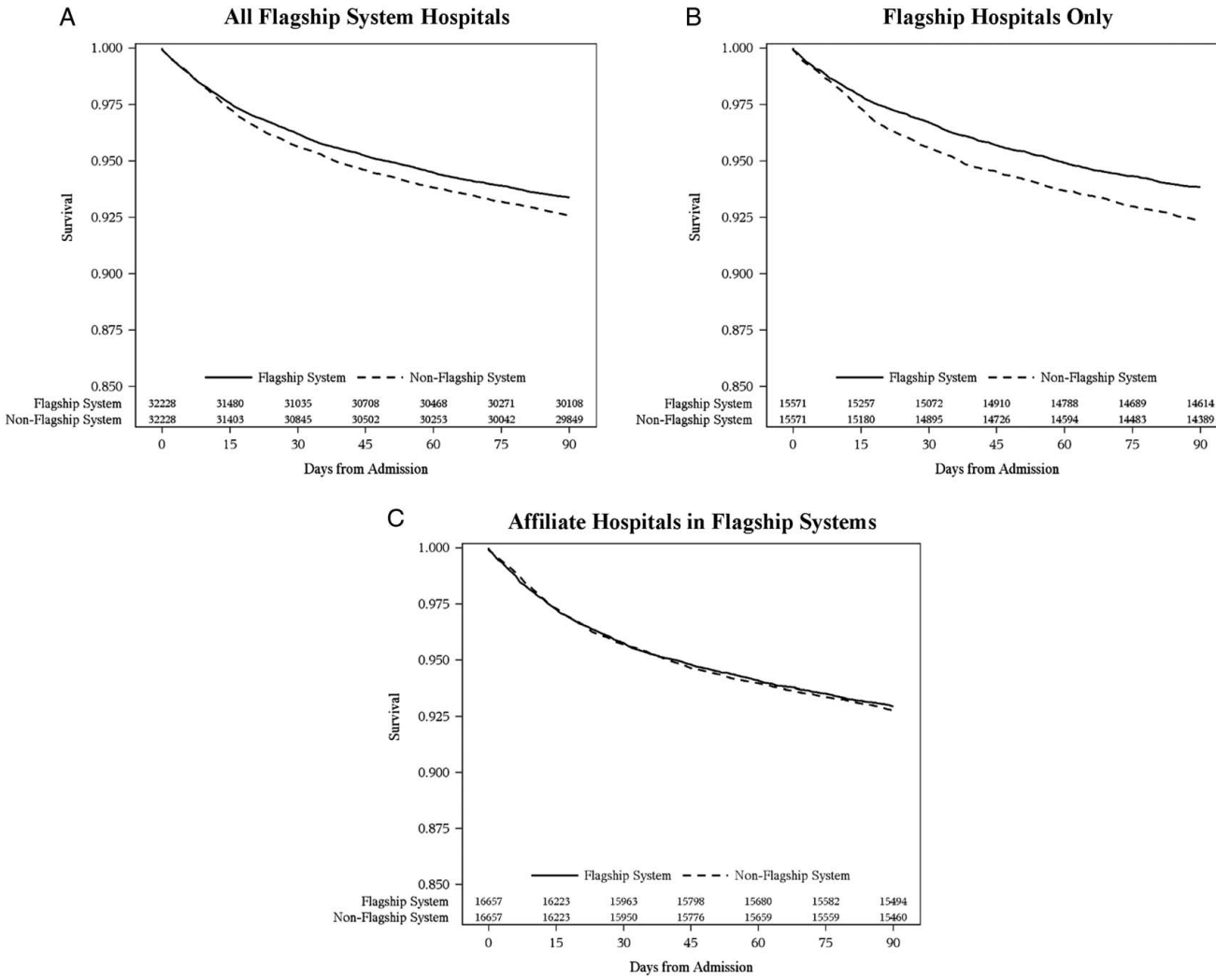

FIGURE 1. Kaplan-Meier survival plots by hospital type versus matched controls outside the flagship system but within the same HRR for (A) all flagship system hospitals, (B) flagship hospitals only, and (C) affiliate hospitals in flagship systems.

DISCUSSION

Hospital systems have often said that their mergers and acquisitions—which have accelerated in recent years[1,14,16,46,47]—are associated with higher quality due to improved care coordination and economies of scale.[1–7] Prior literature has suggested that, despite evidence of variation in surgical outcomes within highly rated hospital systems,[10] many, if not most, patients expect hospitals from the same system to offer the same standard of care, regardless of whether they are at a flagship hospital or its local affiliate.[8]

To study this, we performed a large, carefully matched study across 35 of the largest HRRs, controlling for 32 general surgical procedures, 55 comorbidities and functional status indicators, multimorbidity status, emergent admission status, and sociodemographic variables including age, sex, race, dual-eligibility, and neighborhood education and poverty levels. We found that Medicare patients undergoing inpatient general surgery at flagship system hospitals had lower rates of 30-day and 90-day mortality compared with their matched controls at hospitals outside the flagship system who underwent the same procedure within the same HRR. However, these mortality differences were driven almost entirely by flagship hospitals—the major "brand-name" hospital in each flagship system—and almost entirely by lower mortality for older patients with multimorbidity in those hospitals. By contrast, no mortality difference was observed at affiliate hospitals of flagship systems, and difference-in-differences analysis confirmed that the mortality difference between flagship hospitals and their controls was significantly larger than that between affiliate hospitals and their controls. Similar results were seen for 30-day failure-to-rescue, a surgical quality indicator.[23–26] Therefore, while flagship hospitals exhibited superior surgical outcomes versus their matched controls, affiliates of these flagship hospitals did not.

In addition, we found that older patients with multimorbidity[22] undergoing surgery at flagship hospitals had lower 30-day mortality, 90-day mortality, and 30-day failure-to-rescue rates than matched controls at other hospitals, whereas this was not true for patients without multimorbidity. These disproportionate

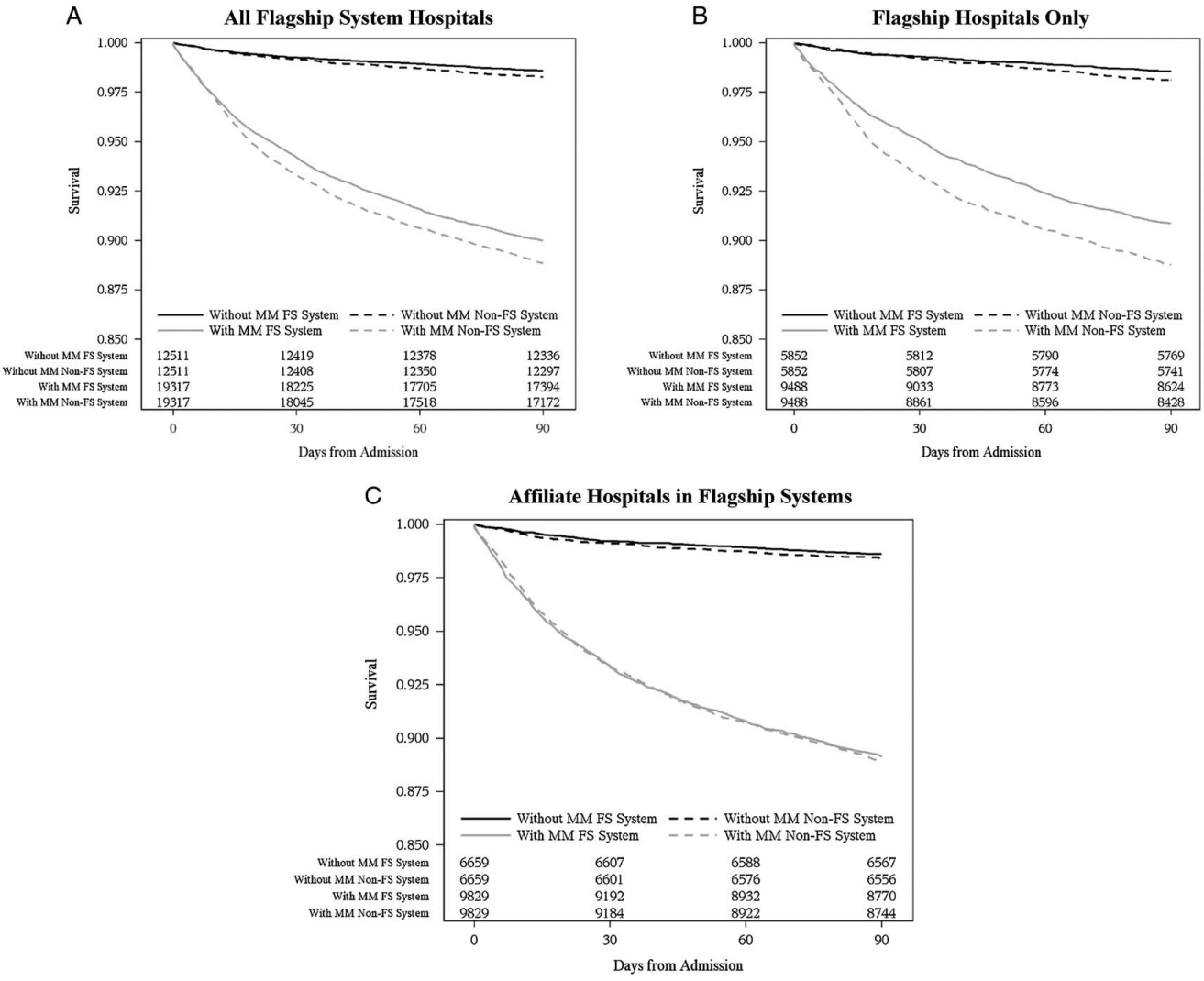

FIGURE 2. Kaplan-Meier survival plots by hospital type and multimorbidity (MM) status versus matched controls outside the flagship system (FS) but within the same HRR for (A) all flagship system hospitals, (B) flagship hospitals only, and (C) affiliate hospitals in flagship systems.

benefits for older patients with multimorbidity at flagship hospitals are consistent with existing literature demonstrating that higher-quality hospitals are associated with superior outcomes for high-risk patients.[21,22,24,37–40] In contrast, we did not find any differential benefits for patients with multimorbidity at affiliate hospitals relative to their controls. We further demonstrated that older patients with multimorbidity appeared to have significantly improved outcomes at flagship hospitals relative to their controls compared with those at affiliate hospitals in the same flagship system relative to their controls.

Our analysis builds on prior work examining surgical outcomes in major hospital systems. Using Medicare data for patients undergoing colectomy, coronary artery bypass graft, or hip replacement in 16 highly rated hospital systems, Sheetz et al[10] uncovered wide variation in surgical outcomes among affiliated hospitals in the same system while also noting that outcomes were not consistently better at flagship versus affiliate hospitals. In contrast, research examining outcomes after complex cancer treatment at top-ranked cancer hospitals versus their affiliates revealed superior survival at the top-ranked hospitals.[11,12] Prior analyses have also suggested that rates of mortality and readmissions for all inpatients do not improve after a hospital is acquired, while patient experience may actually worsen.[13] Our work extends these analyses by comparing flagship hospitals and affiliate hospitals to within-HRR-matched control patients at other hospitals across the breadth of general surgery while also examining whether differential patterns are observed for high-risk patients.

These findings are relevant for patients and policymakers. Patients should not expect superior quality of general surgical care at affiliate hospitals of major regional systems over hospitals outside these systems based solely on their affiliation. In addition, some patients—especially older patients with multimorbidity—may be better served at the flagship hospital itself rather than at its affiliates or hospitals outside the flagship system in the same region.

Our study had limitations. We examined 35 of the largest HRRs in the United States (of 306 HRRs). Each included HRR had a flagship system, including one flagship hospital plus its affiliates

TABLE 4. Rates of 30-Day Failure-to-Rescue for General Surgery Patients in all Flagship System Hospitals, Flagship Hospitals Only, or Affiliated Hospitals in Flagship Systems Compared With Within-HRR-Matched Controls at Other Hospitals

	N	Flagship system hospitals (%)	Matched controls (%)	Difference in rates (%)	95% CI (%)	*P*	Adjusted *P*-value*
Flagship system hospitals versus matched controls							
All patients	32,228	11.12	12.93	–1.81	(–2.65, –0.96)	< 0.001	< 0.001
With multimorbidity	19,317	13.47	15.61	–2.14	(–3.19, –1.08)	< 0.001	< 0.001
Without multimorbidity	12,511	3.63	4.28	–0.65	(–1.69, 0.38)	0.22	1.000
With-versus-without multimorbidity (difference-in-differences)				–1.48	(–2.96, 0.00)	< 0.05	0.398
Flagship hospitals versus matched controls							
All patients	15,571	9.28	12.75	–3.47	(–4.63, –2.32)	< 0.001	< 0.001
With multimorbidity	9488	11.10	15.40	–4.30	(–5.72, –2.89)	< 0.001	< 0.001
Without multimorbidity	5852	3.32	3.94	–0.62	(–2.05, 0.81)	0.39	1.000
With-versus-without multimorbidity (difference-in-differences)				–3.68	(–5.69, –1.67)	< 0.001	0.003
		Affiliate hospitals (%)	Matched controls (%)	Difference in rates (%)	95% CI (%)	*P*	Adjusted *P*-value*
Affiliate hospitals in flagship systems versus matched controls							
All patients	16,657	13.00	13.10	–0.10	(–1.36, 1.16)	0.87	1.000
With multimorbidity	9829	15.91	15.81	0.10	(–1.48, 1.68)	0.90	1.000
Without multimorbidity	6659	3.93	4.60	–0.67	(–2.20, 0.86)	0.39	1.000
With-versus-without multimorbidity (difference-in-differences)				0.77	(–1.43, 2.97)	0.49	1.000
		Flagship difference (%)	Control difference (%)	Difference-in-differences (%)	95% CI (%)	*P*	Adjusted *P*-value*
Difference-in-differences: flagship hospitals minus affiliate hospitals in flagship systems							
Overall		–3.72	–0.35	–3.37	(–5.08, –1.67)	< 0.001	0.001
With multimorbidity		–4.81	–0.41	–4.40	(–6.52, –2.28)	< 0.001	< 0.001
Without multimorbidity		–0.61	–0.66	0.05	(–2.04, 2.14)	0.96	1.000

*Adjusted *P*-values use the Bonferroni-Holm method.[42,43]

and their controls. The hospitals in these 35 HRRs had higher numbers of beds, resident-to-bed ratios, and high technology capabilities compared with the average hospital across the nation, suggesting that the larger HRRs we examined contained better-resourced hospitals than the typical hospital. Also, the study used only fee-for-service Medicare claims. In addition, some information on chronic conditions may be inconsistently recorded across hospitals. We partially addressed this limitation by using a 1-year look-back to obtain information on chronic conditions from both inpatient claims and claims from physician offices and CMS outpatient files, which should have reduced this issue. Further, our definition of multimorbidity[22] and our matched analyses incorporated several forms of objective information, such as functional status indicators obtained from the CMS Durable Medical Equipment files.

In conclusion, we found that while flagship system hospitals offered superior outcomes for patients undergoing inpatient general surgery procedures compared with matched controls outside the flagship system but in the same HRR, those benefits were driven almost entirely by flagship hospitals themselves and concentrated in older patients with multimorbidity. In contrast, affiliates of these flagship hospitals did not offer any significant outcomes benefits over controls. Thus, hospital affiliation with a flagship system does not alone assure better surgical outcomes.

REFERENCES

1. MedPAC. Report to the Congress. Medicare Payment Policy. Chapter 15. Congressional request on health care provider consolidation. Washington, DC: MedPAC; March 2020. Accessed June 28, 2022. https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/mar20_medpac_ch15_sec.pdf
2. Gaynor M, Ho K, Town RJ. The industrial organization of health-care markets. *J Econ Lit*. 2015;53:235–284.
3. Noether M, May S, Stearns B Hospital merger benefits: views from hospital leaders and econometric analysis: an update. September 2019. Accessed June 28, 2022. https://www.aha.org/system/files/media/file/2019/09/cra-report-merger-benefits-2019-f.pdf
4. Dafny LS, Lee TH. The good merger. *N Engl J Med*. 2015;372:2077–2079.
5. Wang E, Arnold S, Jones S, et al. Quality and safety outcomes of a hospital merger following a full integration at a safety net hospital. *JAMA Netw Open*. 2022;5:e2142382.
6. Scanlon DP, Harvey JB, Wolf LJ, et al. Are health systems redesigning how health care is delivered? *Health Serv Res*. 2020;55(suppl 3):1129–1143.
7. Heeringa J, Mutti A, Furukawa MF, et al. Horizontal and vertical integration of health care providers: a framework for understanding various provider organizational structures. *Int J Integr Care*. 2020;20:2.
8. Chiu AS, Resio B, Hoag JR, et al. Why travel for complex cancer surgery? Americans react to "brand-sharing" between specialty cancer hospitals and their affiliates. *Ann Surg Oncol*. 2019;26:732–738.
9. Reames BN, Anaya DA, Are C. Hospital regional network formation and "brand sharing": appearances may be deceiving. *Ann Surg Oncol*. 2019;26:711–713.
10. Sheetz KH, Ibrahim AM, Nathan H, et al. Variation in surgical outcomes across networks of the highest-rated US hospitals. *JAMA Surg*. 2019;154:510–515.
11. Hoag JR, Resio BJ, Monsalve AF, et al. Differential safety between top-ranked cancer hospitals and their affiliates for complex cancer surgery. *JAMA Netw Open*. 2019;2:e191912.
12. Boffa DJ, Mallin K, Herrin J, et al. Survival after cancer treatment at top-ranked US cancer hospitals vs affiliates of top-ranked cancer hospitals. *JAMA Netw Open*. 2020;3:e203942.
13. Beaulieu ND, Dafny LS, Landon BE, et al. Changes in quality of care after hospital mergers and acquisitions. *N Engl J Med*. 2020;382:51–59.

14. Cutler DM, Scott, Morton F. Hospitals, market share, and consolidation. *JAMA*. 2013;310:1964–1970.
15. Xu T, Wu AW, Makary MA. The potential hazards of hospital consolidation: implications for quality, access, and price. *JAMA*. 2015; 314:1337–1338.
16. Mathematica Washington DC, Kimmey L, Machta R, et al. Comparative health system performance initiative: compendium of US Health Systems, 2018. Technical Documentation. AHRQ Publication No. 20 (21)-0011 November 2019 (updated January 2021). Accessed July 12, 2022. https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2018-compendium-techdoc.pdf
17. Wennberg JE, Cooper MM. *The Dartmouth Atlas of Health Care in the United States: The Center for the Evaluative Clinical Sciences*. Chicago, IL: American Hospital Publishing Inc; 1996.
18. Research Data Assistance Center. CMS Virtual Research Data Center (VRDC). 2021. Accessed June 10, 2021. https://www.resdac.org/cms-virtual-research-data-center-vrdc
19. Zhang Y, Baik SH, Fendrick AM, et al. Comparing local and regional variation in health care spending. *N Engl J Med*. 2012;367:1724–1731.
20. Hawkins AT, Samuels LR, Rothman RL, et al. National variation in elective colon resection for diverticular disease. *Ann Surg*. 2022;275:363–370.
21. Silber JH, Reiter JG, Rosenbaum PR, et al. Defining multimorbidity in older surgical patients. *Med Care*. 2018;56:701–710.
22. Ramadan OI, Rosenbaum PR, Reiter JG, et al. Redefining multimorbidity in older surgical patients. *J Am Coll Surg*. 2023;236:1011–1022.
23. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: a study of adverse occurrence and failure to rescue. *Med Care*. 1992;30:615–629.
24. Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and failure to rescue with high-risk surgery. *Med Care*. 2011;49:1076–1081.
25. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: comparing definitions to measure quality of care. *Med Care*. 2007;45:918–925.
26. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009;361:1368–1375.
27. Rosenbaum PR. Optimal matching of an optimally chosen subset in observational studies. *J Comput Graph Stat*. 2012;21:57–71.
28. Rosenbaum PR. Modern algorithms for matching in observational studies. *Annu Rev Stat Appl*. 2020;7:143–176.
29. Fogarty CB, Mikkelsen ME, Gaieski DF, et al. Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *J Am Stat Assoc*. 2016; 111:447–458.
30. Neuman MD, Rosenbaum PR, Ludwig JM, et al. Anesthesia technique, mortality, and length of stay after hip fracture surgery. *JAMA*. 2014;311: 2508–2517.
31. Niknam BA, Zubizarreta JR. Using cardinality matching to design balanced and representative samples for observational studies. *JAMA*. 2022;327:173–174.
32. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39:33–38.
33. Cochran WG, Rubin DB. Controlling bias in observational studies: a review. *Sankhya Ser A*. 1973;35:417–446.
34. Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2:808–840.
35. Silber JH, Rosenbaum PR, McHugh MD, et al. Comparison of the value of nursing work environments in hospitals across different levels of patient risk. *JAMA Surg*. 2016;151:527–536.
36. Silber JH, Rosenbaum PR, Niknam BA, et al. Comparing outcomes and costs of surgical patients treated at major teaching and nonteaching hospitals: a national matched analysis. *Ann Surg*. 2020;271:412–421.
37. Silber JH, Rosenbaum PR, Niknam B, et al. Comparing outcomes and costs of medical patients treated at major teaching and nonteaching hospitals: a national matched analysis. *J Gen Intern Med*. 2020;35: 743–752.
38. Ayanian JZ, Weissman JS. Teaching hospitals and quality of care: a review of the literature. *Milbank Q*. 2002;80:569–593.
39. Burke LG, Frakt AB, Khullar D, et al. Association between teaching status and mortality in US hospitals. *JAMA*. 2017;317:2105–2113.
40. Lasater KB, McHugh MD, Rosenbaum PR, et al. Evaluating the costs and outcomes of hospital nursing resources: a matched cohort study of patients with common medical conditions. *J Gen Intern Med*. 2021;36: 84–91.
41. Fleiss JL, Levin B, Paik MC. Chapter 13. The analysis of data from matched samples. Section 13.1. Matched pairs: dichotomous outcome. In: WA Shewart, SS Wilks, eds. *Statistical Methods for Rates and Proportions*, 3rd ed. New York, NY: John Wiley & Sons; 2003:374–496.
42. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6:65–70.
43. Wright SP. Adjusted *P* values for simultaneous inference. *Biometrics*. 1992;48:1005–1013.
44. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986;1:54–75.
45. SAS Institute. *Version 94 of the Statistical Analytic Software System for UNIX*. Cary, NCSAS Institute Inc. 2013.
46. Furukawa MF, Machta RM, Barrett KA, et al. Landscape of health systems in the United States. *Med Care Res Rev*. 2020;77:357–366.
47. Sherry TB, Damberg CL, DeYoreo M, et al. Is bigger better? A closer look at small health systems in the United States. *Med Care*. 2022;60: 504–511.