

Comparing Outcomes and Costs of Medical Patients Treated at Major Teaching and Non-teaching Hospitals: A National Matched Analysis



Jeffrey H. Silber, MD, PhD^{1,2,3,4,5}, Paul R. Rosenbaum, PhD^{5,6}, Bijan A. Niknam, BS¹, Richard N. Ross, MS¹, Joseph G. Reiter, MS¹, Alexander S. Hill, BS¹, Lauren L. Hochman, BA¹, Sydney E. Brown, MD, PhD³, Alexander F. Arriaga, MD, MPH, ScD^{3,7,8}, and Lee A. Fleisher, MD^{3,5,8}

¹Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ²The Department of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; ³Department of Anesthesiology and Critical Care, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; ⁴Department of Health Care Management, Wharton School, University of Pennsylvania, Philadelphia, PA, USA; ⁵Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, USA; ⁶Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA, USA; ⁷Department of Anesthesiology, Perioperative, and Pain Medicine, Brigham and Women's Hospital, Boston, MA, USA; ⁸Center for Perioperative Outcomes Research and Transformation, University of Pennsylvania, Philadelphia, PA, USA.

BACKGROUND: Teaching hospitals typically pioneer investment in new technology and cultivate workforce characteristics generally associated with better quality, but the value of this extra investment is unclear.

OBJECTIVE: Compare outcomes and costs between major teaching and non-teaching hospitals by closely matching on patient characteristics.

DESIGN: Medicare patients at 339 major teaching hospitals (resident-to-bed (RTB) ratios ≥ 0.25); matched patient controls from 2439 non-teaching hospitals (RTB ratios < 0.05).

PARTICIPANTS: Forty-three thousand nine hundred ninety pairs of patients (one from a major teaching hospital and one from a non-teaching hospital) admitted for acute myocardial infarction (AMI), 84,985 pairs admitted for heart failure (HF), and 74,947 pairs admitted for pneumonia (PNA).

EXPOSURE: Treatment at major teaching hospitals versus non-teaching hospitals.

MAIN MEASURES: Thirty-day all-cause mortality, readmissions, ICU utilization, costs, payments, and value expressed as extra cost for a 1% improvement in survival.

KEY RESULTS: Thirty-day mortality was lower in teaching than non-teaching hospitals (10.7% versus 12.0%, difference = -1.3%, $P < 0.0001$). The paired cost difference (teaching - non-teaching) was \$273 ($P < 0.0001$), yielding \$211 per 1% mortality improvement. For the quintile of pairs with highest risk on admission, mortality differences were larger (24.6% versus 27.6%, difference = -3.0%, $P < 0.0001$), and paired cost difference = \$1289 ($P < 0.0001$), yielding \$427 per 1% mortality improvement at 30 days. Readmissions and ICU utilization were lower in teaching

hospitals (both $P < 0.0001$), but length of stay was longer (5.5 versus 5.1 days, $P < 0.0001$). Finally, individual results for AMI, HF, and PNA showed similar findings as in the combined results.

CONCLUSIONS AND RELEVANCE: Among Medicare patients admitted for common medical conditions, as admission risk of mortality increased, the absolute mortality benefit of treatment at teaching hospitals also increased, though accompanied by marginally higher cost. Major teaching hospitals appear to return good value for the extra resources used.

KEY WORDS: Medicare; value; teaching hospitals; mortality; cost.

J Gen Intern Med

DOI: 10.1007/s11606-019-05449-x

© Society of General Internal Medicine 2019

INTRODUCTION

Studies have previously examined the relationship between hospital teaching status and patient outcomes across a range of medical conditions and surgical procedures.¹⁻⁹ Understanding the value of teaching hospitals is important, because they often invest more than non-teaching hospitals in cutting-edge technology and workforce characteristics generally associated with better quality, but typically, this comes with additional cost.¹ In order to better quantify the impact of this investment in care for medical patients, we use new methods of multivariate matching¹⁰ that closely control for patient differences and compare outcomes and costs on a national scale between teaching and non-teaching hospitals among patients admitted for three common and representative medical conditions that rank among the 10 leading causes of hospitalization (acute myocardial infarction, heart failure, and pneumonia).¹¹ Matching ensures that comparisons between teaching and non-teaching hospitals use pairs of similar patients who have

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-019-05449-x>) contains supplementary material, which is available to authorized users.

Received April 9, 2019

Revised September 15, 2019

Accepted September 26, 2019

both comparable risk levels on admission and also have that risk for comparable reasons, unlike in regression models where two patients can have similar expected risk due to very different comorbidities or risk factors. Through matching, we can assess whether differences in patient outcomes vary significantly within particular subsets of patients or whether outcomes systematically change as patient risk increases, a concept we call risk synergy.^{12–14} By comparing costs and outcomes, we can learn whether specific patient groups fare better at major teaching or non-teaching hospitals and generate essential data points for assessing comparative value (cost per 1% improvement in survival)—crucial for better referral of patients to hospitals and for implementing value-based payments.¹⁵

METHODS

This research study was approved by the Children’s Hospital of Philadelphia Institutional Review Board.

Patient Population

We were granted access via the Centers for Medicare & Medicaid Services (CMS) Virtual Research Data Center (VRDC) to administrative claims data for older Medicare beneficiaries admitted to short-term acute care hospitals in the USA for principal diagnoses of acute myocardial infarction (AMI), heart failure (HF), or pneumonia (PNA) between July 1, 2012, and November 30, 2014. The VRDC hosted each patient’s beneficiary summary file and administrative claims (inpatient, outpatient, carrier/part B, and hospice files). We excluded hospitalizations if the patient was under 65.5 years old at admission, had evidence of hospice care, or lacked fee-for-service Medicare in the 6 months prior to admission, the month of admission, or the month after admission. If a patient had multiple qualifying admissions during the study period, we randomly chose one. After applying these restrictions, we randomly split the sample to use 90% for matching, with the remaining 10% set aside to create baseline prognostic scores.¹⁶ Details of the sample construction, code lists, and sample means are provided in Appendix Tables 1, 2, and 3.

Definitions

Hospital teaching status was defined using resident-to-bed (RTB) ratios from the Medicare Healthcare Cost Report Information System (HCRIS).¹⁷ The focal group of major teaching hospitals was defined as having RTB ratios of 0.25 or greater.^{18,19} The comparison group of non-teaching hospitals was those with RTB ratios below 0.05.

Patient characteristics were defined using the index admission claim and a 6-month look-back in inpatient, outpatient, carrier/part B, and hospice files. We also defined patient age at admission, year of admission, sex, race, emergent admission

type, and encounters at dialysis facilities, skilled nursing facilities, ambulatory surgical centers, and intermediate care facilities, or transfer from another hospital’s inpatient or emergency department immediately preceding admission to the index hospital. Transferred patients’ outcomes were assigned to the first hospital the patient encountered, either as an inpatient or via the emergency department. Comorbidities were defined using methods adapted from CMS.^{20,21} We also defined 5 geographic regions based on the location of the hospital and estimated a propensity score for admission to a major teaching hospital using all the matching covariates.^{22,23}

To calculate patient risk on admission, we fit binary logistic models predicting probability of 30-day death and ICU use with the 10% sample reserved for risk modeling (see Appendix Tables 4, 5, and 6).²⁴ We divided patients in each of the three medical conditions into five quintiles of predicted 30-day mortality. Predicted length of stay and predicted resource-based hospitalization cost were also modeled with the 10% sample using robust regression.^{25–27}

Hospital bed size was defined using the total bed count from HCRIS. We defined each hospital’s nurse-to-bed (NTB) ratio, nurse mix, and technology level using the Medicare Provider of Services file.^{14,28} Comprehensive cardiac technology was defined as the availability of open-heart surgery and the presence of both a coronary care unit and a catheterization lab.^{21,29}

Outcomes

The primary quality outcome was 30-day all-cause, all-location mortality. We also studied all-cause, all-location re-admission or death extending up to 30 days after discharge, with patients who died in the hospital also counted as readmissions. As a stability analysis, we calculated 30-day post-discharge readmission rates in those that survived to discharge. We also examined length of stay and ICU rate of utilization.

We assessed hospital economic performance from the perspective of the health care system using 30-day resource utilization-based cost,^{14,30,31} which represents resources consumed by the health care system in providing care, and acute-care payments tracked by CMS, which represent payments made to the health care system by Medicare, patients, and other primary payers. Resource utilization-based costs track resources through a standardized national price index and do not reflect prices charged to patients or payers. Thirty-day resource costs counted index hospitalization costs plus costs from emergency department, outpatient, or office visits within 30 days of admission, and all costs associated with readmissions that began within 30 days of the index admission date. See Appendix Table 8 for the costing algorithm.

Acute-care payments reflect the amount actually paid for the hospitalization. We focused on two payment calculations. The first was actual payment as tracked in the CMS claims, which includes all payments made by CMS, the beneficiary, and any other third-party payer. The second payment

calculation removes Medicare's adjustments for geography, disproportionate share hospital (DSH), and indirect medical expenditure (IME) payments.³² See Appendix Table 8 for details. As a stability analysis, we repeated all analyses of cost and payment with alternate definitions that added post-acute-care payments.⁵

Defining Value

We defined value (sometimes referred to as incremental cost-effectiveness) as the difference in resource utilization cost between major teaching and non-teaching matched pairs, divided by the difference in 30-day mortality between major teaching and non-teaching matched pairs (or Δ cost/ Δ mortality). Neumann et al.³³ suggest acceptable ratios between \$100 and \$200,000 per life year saved in the USA. We then repeated these value calculations adding post-acute-care payments. We report value as the cost (or payment) increase associated with a 1% improvement in mortality if the ratio's denominator was defined (i.e., significantly different from zero); readers familiar with the cost-per-life-saved metric can obtain it by multiplying our estimate by 100. The 95% confidence interval for the value ratio was derived using the jackknife.³⁴

Statistical Methods

Matching Methodology. Matching of major teaching cases to non-teaching controls was accomplished using the NetFlow procedure in SAS.³⁵ Patients were matched exactly for principal diagnosis category and quintile of 30-day mortality risk. Within exact groups, we used near-exact matching³⁶ with refined balance¹⁰ to control for hospital region. We also used refined balance to match patient demographics, admission characteristics, and risk factors. Subject to the exact and refined balance constraints, we optimally formed the closest sets of pairs between cases and controls by minimizing a Mahalanobis distance³⁷ between cases and controls, which included all comorbidities, demographic, and admission characteristics, the risk scores, and a propensity score for admission to a major teaching hospital (see Appendix Tables 7a–7c). The AMI and HF matches each included 68 covariates, while the pneumonia match included 76 covariates.

Matching was performed first without viewing outcomes.³⁸ We aimed to attain standardized differences in covariate means below 0.1 standard deviations after matching, though below 0.2 is a traditional standard.^{39,40} We also assessed balance using 2-sample randomization tests, specifically the Wilcoxon rank-sum test for each continuous covariate and Fisher's exact test for each binary one, thereby comparing the balance achieved by matching to the balance expected from complete randomization.⁴¹

Comparing Outcomes. Outcomes were compared using paired methods: for binary outcomes, McNemar's test;⁴² for continuous outcomes, M-statistics for matched pairs.^{26,43–45} We further compared outcomes between major teaching

patients and non-teaching controls within each risk quintile. Trends across risk levels were assessed using the Mantel test for trend for binary outcomes⁴⁶ and robust regression for continuous outcomes.^{26,27} Graphs of major-minus-nonteaching outcome differences by risk level were produced using the LOWESS procedure in SAS,⁴⁷ using smoothed pointwise 95% bootstrapped confidence intervals.⁴⁸

RESULTS

Final Patient and Hospital Sample

The sample comprised 203,922 patients treated for AMI, HF, or PNA at 339 major teaching hospitals and 1,049,037 patients treated at 2533 non-teaching hospitals. All major teaching hospitals and their patients were used in the match, while 2439 non-teaching hospitals (96.3%) were represented in the match.

As expected, the structural features of teaching hospitals differed from non-teaching hospitals in many ways, as seen in Table 1, with more teaching hospitals with high-level cardiac technology and higher nurse-to-bed ratios.

The Quality of the Patient Matches

To illustrate the quality of matching, Table 1 displays a few details of the combined AMI, HF, and PNA matches. Not seen in Table 1 are the 5 risk quintiles that were matched exactly within pairs, and the complete list of 26 demographic covariates and 29 comorbid conditions that were matched. Before matching, teaching hospital cases and non-teaching hospital controls differed in many ways. For example, teaching hospital patients were admitted with higher rates of HF, higher rates of chronic kidney disease, and higher rates of diabetes, while non-teaching controls were more likely to be non-Hispanic whites and have higher rates of COPD. After matching, all clinical covariate standardized differences met the balance criteria of 0.10 SDs or less. All covariates within the individual AMI, HF, and PNA matches also met the balance criteria on all covariates. Complete matching tables are displayed in Appendix Tables 9–9c.

Outcome Results

Table 2 reports outcome results for the combined medical match, as well as separately for AMI, HF, and PNA and within the five quintiles of risk. The overall 30-day mortality rate for the 3 medical conditions combined was 10.7% at major teaching hospitals, compared to 12.0% at non-teaching hospitals (difference = -1.3%, $P < 0.0001$). Major teaching hospitals used the ICU less frequently than non-teaching hospitals (19.9% vs. 22.4%, difference = -2.5%, $P < 0.0001$). Length of stay was longer at teaching hospitals (5.5 days vs. 5.1 days, paired difference = 0.4, $P < 0.0001$), but 30-day readmission or death was lower (30.4% vs. 32.3%, difference = -1.9%, $P < 0.0001$). A stability analysis examining 30-day readmission

Table 1 Combined AMI, HF, and PNA Matching Results

Covariates (percent unless noted)		Combined medical									
		Major teaching cases (N = 203,922)	Matched major teaching cases (N = 203,922)	Matched non-teaching controls (N = 203,922)	All non-teaching controls (N = 1,049,037)	Std. Diff.* before match	P value before match	Std. Diff.* after match	P value after match		
Age on admission (years, mean)		80.1	80.1	80.1	80.1	0.00	0.0693	0.00	0.0289		
Age 85+		34.45	34.46	34.46	33.70	0.02	<0.0001	0.00	0.9606		
Sex (% male)		47.74	47.74	48.08	47.47	0.01	0.0248	-0.01	0.0336		
Race (% non-Hispanic white)		72.64	72.64	75.01	85.18	-0.31	<0.0001	-0.06	<0.0001		
Race (% black)		17.61	17.61	16.02	7.14	0.32	<0.0001	0.05	<0.0001		
Emergent admission		87.68	87.68	87.03	75.52	0.32	<0.0001	0.02	<0.0001		
Transfer from other hospital's ED		0.07	0.07	0.07	0.08	0.00	0.1749	0.00	0.7259		
Admission risk of 30-day death, %		12.1	12.1	12.0	12.0	0.01	0.0064	0.00	0.7169		
Principal diagnosis categories											
Acute myocardial infarction											
Anterior STEMI		1.74	1.74	1.74	1.79	0.00	0.1617	0.00	1.0000		
Other location STEMI		2.73	2.73	2.73	2.76	0.00	0.5591	0.00	1.0000		
NSTEMI or unspecified MI		17.09	17.09	17.09	18.06	-0.03	<0.0001	0.00	1.0000		
Pneumonia											
PNA due to anaerobes or Gram-negative bacteria		1.06	1.06	1.06	1.06	0.00	0.8779	0.00	1.0000		
Pneumonia due to <i>H. influenzae</i>		0.09	0.09	0.09	0.12	-0.01	<0.0001	0.00	1.0000		
Legionnaire's disease		0.10	0.10	0.10	0.04	0.02	<0.0001	0.00	1.0000		
Pneumonia due to MRSA		0.56	0.56	0.56	0.77	-0.03	<0.0001	0.00	1.0000		
Pneumonia, other		25.28	25.28	25.28	31.59	-0.14	<0.0001	0.00	1.0000		
Pneumococcal pneumonia		0.38	0.38	0.38	0.41	0.00	<0.0001	0.00	1.0000		
Pneumonia due to <i>Pseudomonas</i>		0.45	0.45	0.45	0.52	-0.01	<0.0001	0.00	1.0000		
Streptococcal pneumonia		0.11	0.11	0.11	0.14	-0.01	0.0126	0.00	1.0000		
Staphylococcal pneumonia		0.04	0.04	0.04	0.05	-0.01	0.0414	0.00	1.0000		
Viral pneumonia		1.37	1.37	1.37	0.89	0.05	<0.0001	0.00	1.0000		
Aspiration pneumonia		7.31	7.31	7.31	7.62	-0.01	<0.0001	0.00	1.0000		
Heart failure											
HF with chronic kidney disease		2.83	2.83	2.83	1.64	0.08	<0.0001	0.00	1.0000		
HF with hypertension		1.11	1.11	1.11	0.87	0.02	<0.0001	0.00	1.0000		
HF, other		37.73	37.73	37.73	31.69	0.13	<0.0001	0.00	1.0000		
Comorbidities											
Hx of hypertension		89.87	89.87	89.28	89.11	0.02	<0.0001	0.02	<0.0001		
Hx of diabetes		48.39	48.39	47.64	45.24	0.06	<0.0001	0.01	<0.0001		
Hx of HF		44.65	44.65	44.68	39.62	0.10	<0.0001	0.00	0.8923		
Hx of COPD		37.49	37.49	37.33	44.50	-0.14	<0.0001	0.00	0.2798		
Hx of renal failure		32.21	32.21	31.41	28.45	0.08	<0.0001	0.02	<0.0001		
Hx of dementia		26.62	26.62	25.87	27.18	-0.01	<0.0001	0.02	<0.0001		
Hx of stroke		9.14	9.14	8.93	7.86	0.05	<0.0001	0.01	0.0224		
Hx of unstable angina		8.13	8.13	7.64	6.64	0.06	<0.0001	0.02	<0.0001		
Hx of cancer		7.13	7.13	6.43	6.01	0.05	<0.0001	0.03	<0.0001		
Hx of AMI		4.47	4.47	4.20	4.96	-0.02	<0.0001	0.01	<0.0001		

(continued on next page)

Table 1. (continued)

Covariates (percent unless noted)	Combined medical		Matched major teaching cases (N = 203,922)	Matched non-teaching controls (N = 203,922)	All non-teaching controls (N = 1,049,037)	Std. Diff.* before match	P value before match	Std. Diff.* after match	P value after match
	Major teaching cases (N = 203,922)	Matched non-teaching controls (N = 203,922)							
Hospital characteristics (not included in the matching algorithm)									
No. of hospitals	339	2439	2533			2.95	<0.0001	2.95	<0.0001
RTB ratio (mean)	0.50	0.00	0.00			—	—	—	—
COTH membership (%)	68.81	0.00	0.00			—	—	—	—
Nurse-to-bed ratio ≥ 1 (%)	85.62	71.21	68.97			0.41	<0.0001	0.35	<0.0001
Cardiac technology (%)	58.60	20.08	21.76			0.81	<0.0001	0.85	<0.0001
Size > 250 beds (%)	85.28	34.94	34.75			1.20	<0.0001	1.20	<0.0001

The table reports balance for selected covariates after matching. All matched covariates met balance criteria. Complete balance tables with all variables, along with individual tables for AMI, HF, and PNA, are provided in Appendix Tables 9a–9c

Hx history

*Standardized difference in units of before-match standard deviation

in patients who survived to discharge displayed similar results, except for the highest severity patients in quantile 5, where teaching hospitals appeared to have higher readmissions than non-teaching hospitals when ignoring differential death rates (see Appendix Table 10). Thirty-day resource utilization costs were \$19,249 versus \$18,944 (a paired difference of \$273, $P < 0.0001$). Thirty-day total acute-care payments were \$22,304 versus \$17,141 (a paired difference of \$5048, $P < 0.0001$); however, after removing geography, DSH, and IME payments, 30-day acute-care payments were \$17,902 versus \$16,207 (a paired difference of \$1662, $P < 0.0001$). See Appendix Tables 11–11c for other forms of payment calculations. Results for the individual medical conditions were similar.

Differences in Outcomes and Costs Across Patient Risk Levels

The last 6 columns of Table 2 report outcome differences for the combined medical group by ascending risk quintiles (risk synergy) and summary tests for trend. We found increasing mortality reductions at major teaching hospitals with increasing risk on admission. There was no difference in mortality in the lowest-risk quintile (2.5% vs. 2.6%, difference = -0.1%, $P = 0.4670$), but a large difference for the highest-risk quintile (24.6% vs. 27.6%, difference = -3.0%, $P < 0.0001$) (trend $P = 0.0001$), and a patient cost difference of \$16,464 versus \$17,000 (paired difference = -\$546, $P < 0.0001$) in the lowest-risk quintile and a patient cost difference of \$22,265 versus \$20,901 (paired difference = \$1289, $P < 0.0001$) in the highest-risk quintile (trend $P < 0.0001$).

Figure 1a, b, and c display the relationship between escalating mean patient risk on admission in matched pairs on the x-axis and the difference in 30-day mortality between the matched pairs on the y-axis, and similarly, Figure 2a, b, and c display the differences in cost. The central curve in each figure shows the paired difference in outcome rate at the risk level indicated on the x-axis, and the shaded region represents the pointwise 95% confidence interval from the bootstrap. Points below zero for the y-axis suggest lower mortality or cost in the teaching hospital patients for a given risk level.

We also saw significant trends (in both directions) in all outcomes between major teaching and non-teaching patients, except for length of stay and ICU use. Notably, teaching hospitals consistently used the ICU at lower rates than non-teaching hospitals.

Estimates of Comparative Value

Table 2 also provides estimates of the difference in cost for a 1% difference in mortality between major teaching and non-teaching hospitals, with associated 95% CIs. We also examine value within risk quintile and again provide condition-specific value estimates in Appendix Tables 12–12c.

Tracking price-standardized resource utilization costs, the value estimate for combined medical conditions was \$211 for

Table 2 Combined Acute Myocardial Infarction, Heart Failure, and Pneumonia Outcomes, Costs, and Value by Hospital Type and Patient Risk on Admission

Outcome (% unless noted)	Cohort	Combined quintiles of risk on admission (Q1 = lowest)					Trend P value*				
		AMI (N = 43,990 pairs)	HF (N = 84,985 pairs)	PNA (N = 74,947 pairs)	Combined patients (N = 203,922 pairs)	Q1 (N = 41,851 pairs)		Q2 (N = 39,207 pairs)	Q3 (N = 39,738 pairs)	Q4 (N = 40,547 pairs)	Q5 (N = 42,579 pairs)
30-day mortality (%)	% of pairs	21.5	41.7	36.8	100.0	20.5	19.2	19.5	19.9	20.9	0.0001
	Teaching	12.1	9.3	12.7	10.7	2.5	4.8	7.9	13.1	24.6	
	Non-teaching	12.9	8.3	14.5	12.0	2.6	5.4	8.9	14.9	27.6	
	Difference	-0.9 ^d	-1.0 ^d	-1.8 ^d	-1.3 ^d	-0.1	-0.6 ^c	-0.9 ^d	-1.8 ^d	-3.0 ^d	
30-day resource cost (\$, m-est.)	Teaching	23,875	19,067	16,762	19,249	16,464	18,378	19,041	20,051	22,265	<0.0001
	Non-teaching	23,327	18,267	17,143	18,944	17,000	18,489	18,848	19,450	20,901	
	Difference	489 ^d	756 ^d	-382 ^d	273 ^d	-546 ^d	-140	182	557 ^d	1289 ^d	
	Value estimate (95% CI)	563 (138, 989)	730 (465, 994)	-208 (-300, -116)	211 (119, 303)	N.C. (N.C.)	-234 (-680, 211)	193 (-102, 489)	318 (143, 493)	427 (309, 546)	
30-day total payments (\$, m-est.)	Teaching	29,930	21,017	19,314	22,304	21,075	22,125	21,863	22,490	23,913	<0.0001
	Non-teaching	24,060	15,677	14,789	17,141	16,703	17,420	17,044	16,899	17,624	
	Difference	5795 ^d	5026 ^d	4435 ^d	5048 ^d	4277 ^d	4629 ^d	4715 ^d	5631 ^d	6120 ^d	
	Value estimate (95% CI)	6674 (3488, 9860)	5022 (3729, 6315)	2419 (1968, 2870)	3904 (3336, 4472)	N.C. (N.C.)	7755 (3818, 11,692)	5010 (2929, 7091)	3097 (2257, 3938)	2030 (1634, 2425)	
30-day payments without Geo, IIME, or DSH (\$, m-est.)	Teaching	23,784	16,984	15,503	17,902	16,855	17,786	17,587	18,045	19,197	<0.0001
	Non-teaching	22,437	14,875	14,073	16,207	15,746	16,431	16,145	16,001	16,671	
	Difference	1332 ^d	2054 ^d	1408 ^d	1662 ^d	1086 ^d	1349 ^d	1420 ^d	1972 ^d	2440 ^d	
	Value estimate (95% CI)	1534 (760, 2308)	1981 (1451, 2511)	768 (612, 925)	1286 (1080, 1492)	N.C. (N.C.)	2259 (990, 3529)	1509 (832, 2185)	1125 (796, 1453)	809 (641, 977)	
ICU use (%)	Teaching	46.8	13.8	11.1	19.9	16.6	19.6	19.8	20.7	22.7	0.0921
	Non-teaching	48.7	16.1	14.0	22.4	18.6	22.3	22.5	23.0	25.3	
	Difference	-2.0 ^d	-2.3 ^d	-2.9 ^d	-2.5 ^d	-2.0 ^d	-2.7 ^d	-2.7 ^d	-2.3 ^d	-2.6 ^d	
30-day readmission or death (%)	Teaching	32.9	29.5	30.0	30.4	20.9	25.0	28.2	33.3	44.1	0.0003
	Non-teaching	36.2	30.9	31.6	32.3	23.6	27.1	29.9	34.8	45.5	
	Difference	-3.3 ^d	-1.4 ^d	-1.6 ^d	-1.9 ^d	-2.8 ^d	-2.1 ^d	-1.7 ^d	-1.5 ^d	-1.4 ^d	
Length of stay (days, m-est.)	Teaching	5.3	5.5	5.7	5.5	4.5	5.0	5.4	5.9	6.8	0.6208
	Non-teaching	4.6	4.9	5.6	5.1	4.1	4.6	4.9	5.4	6.3	
	Difference	0.7 ^d	0.6 ^d	0.1 ^d	0.4 ^d	0.4 ^d	0.4 ^d	0.4 ^d	0.4 ^d	0.5 ^d	

Value expressed as increased cost or payment at teaching hospitals per 1% reduction in mortality. Individual tables for AMI, HF, and PNA are provided in the Appendix. P value key is as follows: ^aP < 0.05, ^bP < 0.01, ^cP < 0.001, and ^dP < 0.0001. Non-significant differences have no letter. P values for binary overall outcomes and within strata were calculated using McNemar's test. P values and estimates for continuous overall outcomes and within strata were calculated using M-statistics. Value is well defined if there is a significant difference in mortality between groups. Otherwise in this table, values are reported as N.C. (not calculated). For all value estimates and confidence intervals reported, all 1000 jackknives concur in finding a statistically significant difference in mortality

*Trends for binary outcomes were tested using the Mantel test for trend, while trends for continuous outcomes were tested using robust regression

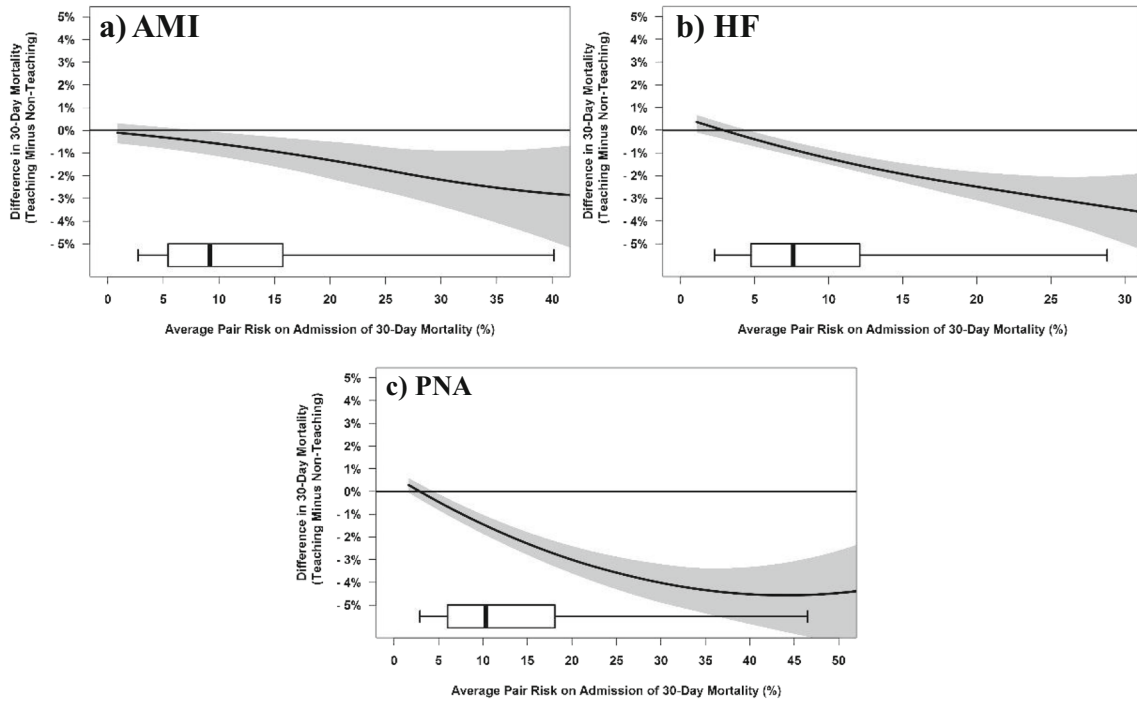


Figure 1 AMI (a), HF (b), and PNA (c) Risk Synergy LOWESS plots of difference in mortality between matched pairs (teaching – non-teaching) versus risk of 30-day mortality on admission. The central curve shows the paired difference in outcome rate as risk level escalates, and the shaded region represents the pointwise 95% confidence interval from the bootstrap. Below each graph is a boxplot of the distribution by a matched pair of average risk of mortality on admission, with whiskers extending to the 5th and 95th percentiles.

a 1% reduction in mortality at the major teaching hospitals compared to non-teaching hospitals, or equivalently \$21,100 per life saved. Similar results were seen within the individual medical conditions. For the lowest-risk patients, costs were

significantly lower at major teaching hospitals, but mortality was not significantly different between hospital types. In the highest-risk quintile, we found a value estimate of \$427 per 1% reduction in mortality. A stability analysis incorporating

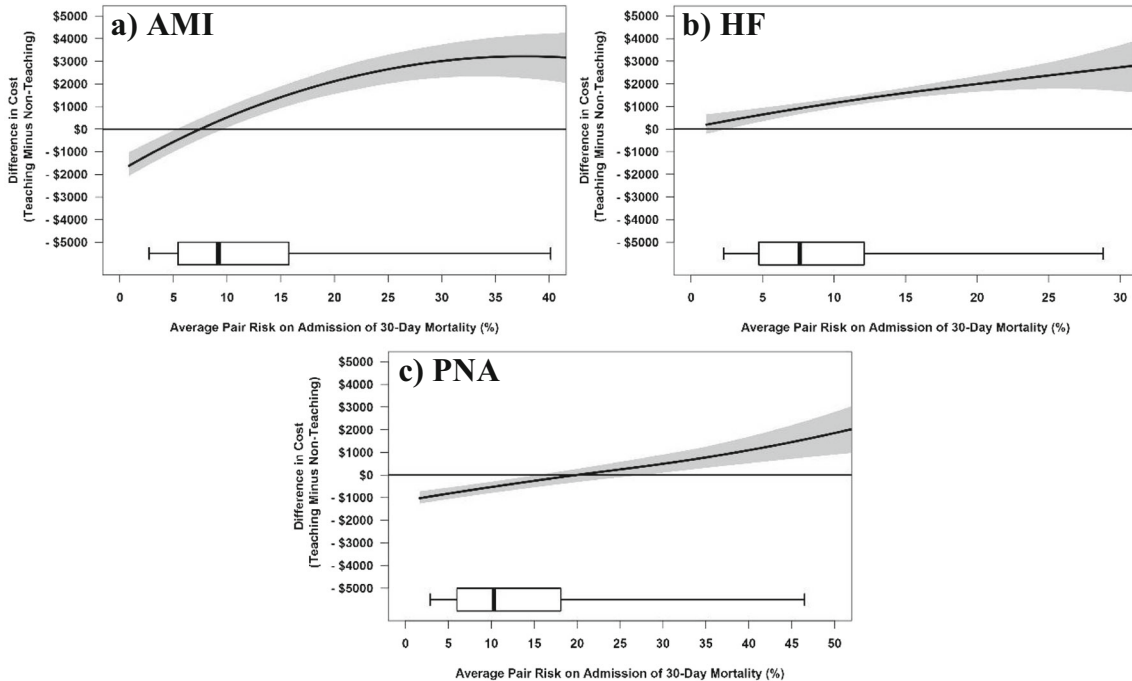


Figure 2 AMI (a), HF (b), and PNA (c) Risk Synergy LOWESS plots of difference in resource utilization (cost) between matched pairs (teaching – non-teaching) versus risk of 30-day mortality on admission. The central curve shows the paired difference in outcome rate as risk level escalates, and the shaded region represents the pointwise 95% confidence interval from the bootstrap. Below each graph is a boxplot of the distribution by a matched pair of average risk of mortality on admission, with whiskers extending to the 5th and 95th percentiles.

post-acute-care costs from all sources displayed very similar results (see Appendix Table 13 for both costs and payments).

Examining total acute-care payments, the overall payment difference for the combined medical conditions was \$5048 with an associated mortality reduction of -1.3% , resulting in an incremental cost-effectiveness ratio point estimate of \$3904 per 1% reduction in mortality. However, removing Medicare payment adjustments for geography and DSH (which would be paid to any hospital in the same location serving the same population) and removing the indirect medical education adjustment (because payments to support medical education are needed to supply future practicing physicians for both teaching and non-teaching hospitals), the payment difference was \$1662, translating to a value estimate of \$1286 per 1% reduction in mortality. For the highest-risk patients, the value estimate using total payments was \$2030. However, after removing geographic, IME, and DSH payments, the highest-risk quintile had a value estimate of \$809.

CONCLUSIONS

Risk Synergy is the phenomenon where increasing patient risk on admission interacts with hospital type¹²⁻¹⁴ to influence outcomes. In this study, as patients presented with poorer health, they experienced increasing 30-day mortality benefit of treatment at major teaching hospitals relative to non-teaching hospitals. A similar finding was observed in our earlier work concerning different nursing environments¹⁴ and also was recently reported in a study by Burke et al.⁷ The present study builds on the work of Burke et al.^{5,7,8} by examining concomitant costs, and value, showing increasing differences in resource utilization costs between teaching and non-teaching hospitals as risk increased, co-occurring with increasing differences in mortality and better value. Major teaching hospitals required fewer resources to treat low-risk patients than non-teaching hospitals did, but as patient risk increased, resource costs at major teaching hospitals increased above those in non-teaching hospitals. In general, however, costs were higher at major teaching hospitals, in part because length of stay was consistently longer across risk levels, and because of the higher resource utilization at teaching hospitals in the highest risk patients compared to non-teaching hospitals.

For the lowest-risk patients in the combined analysis, mortality was not different between hospital types, but cost was typically lower at teaching hospitals. In the second-lowest risk quintile, both mortality and cost were significantly lower at teaching hospitals—a clear indication of accruing value at the teaching hospitals. From the middle quintile to the highest-risk quintile, we observed significantly lower mortality at the teaching hospitals, but also higher costs. By the highest quintile of risk, there was a -3.0% absolute difference in mortality at major teaching hospitals, with an associated \$1289 cost

increase. In our value calculation, that translates to about \$427 per 1% increase in survival.

While a more traditional value calculation of differences in lives saved is useful to policymakers who want to know the statistical cost for one life saved over a population, the information should be interpreted differently by patients. Patients cannot pay \$42,700 for a 100% reduction in their risk of death. Rather, the individual patient may want to think about value as the incremental cost for, say, a 1% absolute reduction in the risk of death.

For both teaching and non-teaching hospitals, it appears that as patient risk increased, resource utilization (cost) increased, but payments were relatively stable. This could represent a cross-subsidization that occurs between low- and high-risk patients at both types of hospitals.

Overall, we found a -1.3% absolute reduction in mortality at major teaching hospitals (10.7% vs. 12.0%) compared to non-teaching hospitals, consistent with a report by Burke et al.,⁸ who found a similar difference in 30-day mortality in national Medicare data with regression modeling. Our analysis also examined resource costs and payments, and the cost-to-quality trade-off along the continuum of patient risk. In so doing, we found that even though resource costs and payments increased at teaching hospitals relative to non-teaching hospitals as patient risk increased, the concomitantly increasing quality benefit associated with lower mortality resulted in good value. While other studies have compared quality and cost between teaching and non-teaching hospitals,^{3,7-9} all such studies fundamentally depend on the specifications of the regression model to make an unbiased comparison between hospital groups whose patients may differ substantially in characteristics and risk. Using matching, we could assure that risk subsets of our analysis compared similar patients with similar diagnoses, thereby allowing us to better understand differential outcomes across and within levels of risk.

Recent work by Burke et al.,^{5,7,8} and summarized in an editorial on value by Khullar et al.,⁶ did not actually estimate value between hospital types. Our study estimated value differences between teaching and non-teaching hospitals, by summarizing the cost differences across a group of matched pairs, and the mortality differences across the same patients included in the matched pairs, and calculated value, again using those same patients. The methods we described also allowed us to place a confidence interval around all of these quantities.

Our study has implications for practice and policy. Treatment capacity in major teaching hospitals is limited, and our study shows that although complex, high-risk patients are currently treated at both major and non-teaching hospitals, major teaching hospitals provide superior outcomes for such patients at only a slight fiscal premium, providing good value. The health system could use this information to promote a more optimal, efficient allocation of patients to academic hospitals, based on both their own needs and the differential capacity of facilities to address them. We also now provide a

quality context to differential payments, which is important knowledge for setting reimbursement policy.

Our study is based on Medicare claims, with all of the limitations associated with such administrative data, including the possibility of confounders not measured in claims. Administrative data also lacks information on quality of life or non-medical costs to patients that would have improved comparisons of cost-effectiveness.

In brief, for comparable patients, 30-day mortality was lower at teaching hospitals for AMI, HF, and PNA, with larger differences for high-risk patients. The cost of treating lower-risk patients was lower at teaching hospitals, while higher-risk patients accrued higher costs at teaching hospitals but had significantly better outcomes, suggesting that extra resources used at major teaching versus non-teaching hospitals represented good value.

Acknowledgments: We thank Traci Frank, AA; Kathryn Yucha, MSN, RN; and Sujatha Changolkar (Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, PA, USA) for their assistance with this research.

Corresponding Author: Jeffrey H. Silber, MD, PhD; Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA (e-mail: silber@email.chop.edu).

Compliance with Ethical Standards:

Conflict of Interest: This research was funded by a grant from the Association of American Medical Colleges (AAMC) to study differences between teaching and non-teaching hospitals on outcomes, costs and value. AAMC had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or decision to submit the manuscript for publication.

The authors declare that the Children's Hospital of Philadelphia (CHOP) and the University of Pennsylvania (PENN) received a research grant from the Association of American Medical Colleges (AAMC) which, in turn, provided partial salary support for some investigators.

REFERENCES

1. Taylor DH, Whellan DJ, Sloan FA. Effects of admission to a teaching hospital on the cost and quality of care for Medicare beneficiaries. *N Engl J Med*. 1999;340:293–99.
2. Allison JJ, Kiefe CI, Weissman NW, et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. *JAMA*. 2000;284:1256–62.
3. Rosenthal GE, Harper DL, Quinn LM, Cooper GS. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA*. 1997;278:485–90.
4. Navathe A, Silber JH, Zhu J, Volpp KE. Does admission to a teaching hospital affect acute myocardial infarction survival? *Acad Med*. 2013;88:475–82.
5. Burke LG, Khullar D, Zheng J, Frakt AB, Orav EJ, Jha AK. Comparison of costs of care for Medicare patients hospitalized in teaching and nonteaching hospitals. *JAMA Netw Open*. 2019;2:e195229.
6. Khullar D, Frakt AB, Burke LG. Advancing the academic medical center value debate. Are teaching hospitals worth it? *JAMA*. 2019;322:205–06.
7. Burke L, Khullar D, Orav EJ, Zheng J, Frakt A, Jha AK. Do academic medical centers disproportionately benefit the sickest patients? *Health Aff (Millwood)*. 2018;37:864–72.
8. Burke LG, Frakt AB, Khullar D, Orav EJ, Jha AK. Association between teaching status and mortality in US hospitals. *JAMA*. 2017;317:2105–13.
9. Ayanian JZ, Weissman JS, Chasan-Taber S, Epstein AM. Quality of care for two common illnesses in teaching and nonteaching hospitals. *Health Aff*. 1998;17:194–205.
10. Pimentel SD, Kelz RR, Silber JH, Rosenbaum PR. Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *J Am Stat Assoc*. 2015;110:515–27.
11. Healthcare Cost and Utilization Project (HCUP). HCUP Fast Stats—Most Common Diagnoses for Inpatient Stays Rockville, MD: Agency for Healthcare Research and Quality, 2019. Available at: <https://www.hcup-us.ahrq.gov/faststats/national/inpatientcommondiagnoses.jsp>. Accessed July 17, 2019.
12. Silber JH, Rosenbaum PR, Ross RN, et al. Indirect standardization matching: Assessing specific advantage and risk synergy. *Health Serv Res*. 2016;51:2330–57.
13. Silber JH, Rosenbaum PR, Wang W, et al. Auditing practice style variation in pediatric inpatient asthma care. *JAMA Pediatr*. 2016;170:878–86.
14. 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;151:527–36.
15. Burwell SM. Setting value-based payment goals—HHS efforts to improve U.S. health care. *N Engl J Med*. 2015;372:897–9.
16. Hansen BB. The prognostic analogue of the propensity score. *Biometrika*. 2008;95:481–88.
17. CMS.gov. Cost Reports 2017. Available at: <https://www.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/cost-reports/>. Accessed June 9, 2017.
18. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. *JAMA*. 2007;298:975–83.
19. Ayanian JZ, Weissman JS. Teaching hospitals and quality of care: A review of the literature. *Milbank Q*. 2002;80:569–93.
20. Yale New Haven Health Services Corporation- Center for Outcomes Research and Evaluation. 2016 Condition-Specific Measures Updates and Specifications Report: Hospital-Level 30-day Risk-Standardized Mortality Measures. Acute Myocardial Infarction. Version 10.0. 2016. Available at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html>. Accessed February 7, 2017.
21. Niknam BA, Arriaga AF, Rosenbaum PR, et al. Adjustment for atherosclerosis diagnosis distorts the effects of percutaneous coronary intervention and the ranking of hospital performance. *J Am Heart Assoc*. 2018;7:pii: e008366.
22. Rosenbaum PR. Chapter 5: Between Observational Studies and Experimentation. In: *Observation and Experiment An Introduction to Causal Inference*. Cambridge: Harvard University Press; 2017:90–96.
23. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
24. SAS Institute. Chapter 53: The LOGISTIC Procedure. In: *SAS/STAT 93 User's Guide*. Cary: SAS Institute, Inc.; 2011:4033–267. Available at: <https://support.sas.com/documentation/cdl/en/statug/63962/PDF/default/statug.pdf>. Accessed February 29, 2008.
25. SAS Institute. Chapter 77: The ROBUSTREG Procedure. In: *SAS/STAT 93 User's Guide*. Cary: SAS Institute Inc.; 2011:6531–625. Available at: <https://support.sas.com/documentation/cdl/en/statug/63962/PDF/default/statug.pdf>. Accessed August 6, 2018.
26. Huber PJ. *Robust Statistics*. Hoboken: Wiley; 1981.
27. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Chapter 6. Linear Models: Robust Estimation. Section 6.3. *M-Estimators for Linear Models*. In: *Robust Statistics The Approach Based on Influence Functions*. 2nd ed. New York: Wiley; 1986:315–28.
28. Centers for Medicare & Medicaid Services. Provider of Services Current Files Baltimore, MD: 2017. Available at: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/>. Accessed June 9, 2017.
29. Silber JH, Arriaga AF, Niknam BA, Hill AS, Ross RN, Romano PS. Failure-to-rescue after acute myocardial infarction. *Med Care*. 2018;56:416–23.
30. Silber JH, Rosenbaum PR, Kelz RR, et al. Medical and financial risks associated with surgery in the elderly obese. *Ann Surg*. 2012;256:79–86.
31. Halpern NA, Pastores SM. Critical care medicine in the United States 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit Care Med*. 2010;38:65–71.
32. Krinsky S, Ryan AM, Mijanovich T, Blustein J. Variation in payment rates under Medicare's inpatient prospective payment system. *Health Serv Res*. 2017;52:676–96.

33. **Neumann PJ, Cohen JT, Weinstein MC.** Updating cost-effectiveness—the curious resilience of the \$50,000-per-QALY threshold. *N Engl J Med.* 2014;371:796–7.
34. **Efron B, Gong G.** A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat.* 1983;37:36–48.
35. SAS Institute. Version 9.4 of the Statistical Analytic Software System for UNIX. Cary: SAS Institute, Inc.; 2013.
36. **Rosenbaum PR.** Chapter 9: Various Practical Issues in Matching. 9.2 Almost Exact Matching. In: *Design of Observational Studies.* New York: Springer; 2010:190–92.
37. **Rubin DB.** Bias reduction using Mahalanobis metric matching. *Biometrics.* 1980;36:293–98.
38. **Rubin DB.** The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat Med.* 2007;26:20–36.
39. **Rosenbaum PR, Rubin DB.** Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat.* 1985;39:33–38.
40. **Cochran WG, Rubin DB.** Controlling bias in observational studies. A review. *Sankhya, Series A.* 1973;35:417–46.
41. **Rosenbaum PR.** Chapter 9.1: Various Practical Issues in Matching: Checking Covariate Balance. In: *Design of Observational Studies.* New York: Springer; 2010:187–90.
42. **Bishop YMM, Fienberg SE, Holland PW.** *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: The MIT Press; 1975.
43. **Rosenbaum PR.** Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. *Biometrics.* 2007;63:456–64. (R package sensitivitymv and sensitivitymw).
44. **Maritz JS.** A note on exact robust confidence intervals for location. *Biometrika.* 1979;66:163–66.
45. **Rosenbaum PR.** Two R packages for sensitivity analysis in observational studies. *Obs Stud.* 2015;1:1–17.
46. **Mantel N.** Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc.* 1963;58:690–700.
47. **Cleveland WS.** Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc.* 1979;74:829–36.
48. **Efron B, Tibshirani R.** Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci.* 1986;1:54–75.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.