

Original Investigation

Comparison of the Value of Nursing Work Environments in Hospitals Across Different Levels of Patient Risk

Jeffrey H. Silber, MD, PhD; Paul R. Rosenbaum, PhD; Matthew D. McHugh, PhD, JD, RN, MPH; Justin M. Ludwig, MA; Herbert L. Smith, PhD; Bijan A. Niknam, BS; Orit Even-Shoshan, MS; Lee A. Fleisher, MD; Rachel R. Kelz, MD, MSCE; Linda H. Aiken, PhD, RN

IMPORTANCE The literature suggests that hospitals with better nursing work environments provide better quality of care. Less is known about value (cost vs quality).

OBJECTIVES To test whether hospitals with better nursing work environments displayed better value than those with worse nursing environments and to determine patient risk groups associated with the greatest value.

DESIGN, SETTING, AND PARTICIPANTS A retrospective matched-cohort design, comparing the outcomes and cost of patients at focal hospitals recognized nationally as having good nurse working environments and nurse-to-bed ratios of 1 or greater with patients at control group hospitals without such recognition and with nurse-to-bed ratios less than 1. This study included 25 752 elderly Medicare general surgery patients treated at focal hospitals and 62 882 patients treated at control hospitals during 2004-2006 in Illinois, New York, and Texas. The study was conducted between January 1, 2004, and November 30, 2006; this analysis was conducted from April to August 2015.

EXPOSURES Focal vs control hospitals (better vs worse nursing environment).

MAIN OUTCOMES AND MEASURES Thirty-day mortality and costs reflecting resource utilization.

RESULTS This study was conducted at 35 focal hospitals (mean nurse-to-bed ratio, 1.51) and 293 control hospitals (mean nurse-to-bed ratio, 0.69). Focal hospitals were larger and more teaching and technology intensive than control hospitals. Thirty-day mortality in focal hospitals was 4.8% vs 5.8% in control hospitals ($P < .001$), while the cost per patient was similar: the focal-control was $-\$163$ (95% CI = $-\$542$ to $\$215$; $P = .40$), suggesting better value in the focal group. For the focal vs control hospitals, the greatest mortality benefit (17.3% vs 19.9%; $P < .001$) occurred in patients in the highest risk quintile, with a nonsignificant cost difference of $\$941$ per patient ($\$53\,701$ vs $\$52\,760$; $P = .25$). The greatest difference in value between focal and control hospitals appeared in patients in the second-highest risk quintile, with mortality of 4.2% vs 5.8% ($P < .001$), with a nonsignificant cost difference of $-\$862$ ($\$33\,513$ vs $\$34\,375$; $P = .12$).

CONCLUSIONS AND RELEVANCE Hospitals with better nursing environments and above-average staffing levels were associated with better value (lower mortality with similar costs) compared with hospitals without nursing environment recognition and with below-average staffing, especially for higher-risk patients. These results do not suggest that improving any specific hospital's nursing environment will necessarily improve its value, but they do show that patients undergoing general surgery at hospitals with better nursing environments generally receive care of higher value.

JAMA Surg. doi:10.1001/jamasurg.2015.4908
Published online January 20, 2016.

 [Invited Commentary](#)

 [Supplemental content at jamasurgery.com](#)

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Jeffrey H. Silber, MD, PhD, Center for Outcomes Research, Children's Hospital of Philadelphia, 3535 Market St, Ste 1029, Philadelphia, PA 19104 (silberj@wharton.upenn.edu).

Past studies have shown that hospitals with excellent nursing environments, as confirmed in a national peer-assessed recognition program, have lower mortality¹⁻⁵ and lower failure-to-rescue rates,⁶ yet others have reported unclear patient cost and revenue benefits associated with hospitals known to have good nursing work environments.^{7,8}

This study asks whether selecting hospitals based solely on excellent nursing environments (defined by having both national peer-assessed recognition and above-average nurse staffing) identifies a set of hospitals that display better outcomes and value, a question most relevant to a patient seeking advice on where to go for care. Our approach was different from previous studies. We did not ask whether a specific hospital would benefit from improving its nursing environment, a question relevant to an administrator capable of changing the environment of the hospital. Therefore, we purposely did not match on individual hospital characteristics, instead seeking to compare 2 groups of hospitals with very different nursing environments but very similar patients and allowing other hospital characteristics to vary naturally with the 2 groups.

Furthermore, by closely matching pairs of patients from hospitals with better and worse nursing environments, we explored whether better nursing environments especially benefit patients of higher initial risk.

Methods

Study Population

The data set comprised Medicare fee-for-service claims for elderly patients admitted for general surgery in Illinois, New York, and Texas from 2004-2006. We acquired the following files: the Master Beneficiary Summary File, inpatient claims, outpatient claims, and Carrier/Part B bills. The study was conducted between January 1, 2004, and November 30, 2006; this analysis was conducted from April to August 2015. This research protocol was reviewed by the Children's Hospital of Philadelphia institutional review board and judged to be non-human subjects research, and informed consent was not required.

Patient Characteristics

Patient characteristics were defined using the index admission and a 90-day look-back in all utilization files. Variables included patient age, year of admission, sex, race, emergency department admission status, transfer-in status, and 31 comorbidities (eAppendix 1 in the [Supplement](#)). Patient probability of 30-day death was estimated by a model fit to an external data set that was not used for matching (eAppendix 2 in the [Supplement](#)). We estimated a propensity score using all of the matching covariates^{9,10} for attending a hospital with a good work environment. We also required an exact match within pairs on all 4-digit principal procedure codes (N = 130; see eAppendix 3 in the [Supplement](#) for the complete list).

Hospital Characteristics

We defined each hospital's nursing environment using the 2007 list of a national voluntary accreditation program for nursing environment excellence that has been found by many stud-

ies to identify hospitals with significantly better nursing environments.^{2,11} Each hospital's nurse-to-bed (NTB) ratio, resident-to-bed ratio, nurse mix, technology level, and number of beds were determined using the Medicare Provider of Service file. The NTB ratio was defined by dividing the number of full-time-equivalent registered nurses and licensed practical nurses by the number of total beds. Likewise, the resident-to-bed ratio was defined by dividing the number of residents by the number of total beds. Nurse mix was the proportion of registered nurses among the total number of registered nurses and licensed practical nurses. Technology level was considered high by the presence of a burn unit or the provision of coronary artery bypass graft surgery or organ transplantation.^{12,13}

Outcomes

Thirty-day mortality was our primary quality-related outcome. We also report in-hospital mortality, in-hospital and 30-day complications (38 common complications that occur after surgery, as defined in previous work; eAppendix 4 in the [Supplement](#)),¹⁴⁻¹⁷ in-hospital and 30-day failure to rescue,^{14,18} all-cause readmissions within 30 days of discharge, length of stay, and intensive care unit (ICU) use.

We used 2 approaches to assess economic performance: costs and Centers for Medicare and Medicaid Services (CMS) payments. Our primary metric was 30-day cost. We calculated each patient's in-hospital and in-hospital plus 30-day costs (hereafter referred to as 30-day costs) based on resource utilization.^{15,19} As in previous studies,^{15,16} in-hospital costs accounted for any resources used for the patient's care during the period of the index hospitalization. Thirty-day costs included in-hospital costs, plus any emergency department, outpatient visit, or office visit costs, as well as any costs arising from a rehospitalization that began within 30 days of the index admission date (counting all costs from the entirety of the readmission, including beyond 30 days). Our costing function was based on data available in Medicare claims. Cost was a function of days in the hospital and level of care (ICU vs floor) for each day, total relative value units determined from all bills, all procedures for which a bill was identified and charged to CMS (including operating room cost and anesthesia), and any bill observed using the description provided here. Finally, we added an estimate of costs directly associated with above- or below-average NTB ratio. The costing algorithm used salary data from the Bureau of Labor and Statistics, adjusted for fringe benefits²⁰ to create adjusted costs reflecting the hospital's positive or negative deviation from the average NTB ratio, and assigned to each patient an additional cost or cost reduction reflecting the extra or reduced nursing costs per day multiplied by days spent on the general floor (eAppendix 5 in the [Supplement](#)). We also report a second cost metric (cost without NTB adjustment) that did not include adjustments to cost based on differing NTB ratios.

Another approach to evaluating value was through Medicare payments associated with the hospital admission (eAppendix 5 and eAppendix 6 in the [Supplement](#)). We report payments using 2 definitions. One included all the payments provided by CMS. A second definition omitted the geography adjustment (because we did not want possibly different pric-

Table 1. Hospital Characteristics

Characteristic ^a	Hospitals		P Value ^b
	Focal (n = 35)	Control (n = 293)	
External validation of excellent nursing work environment, %	100.0	0	<.001
NTB ratio, mean	1.51	0.69	<.001
Nurse skill mix, mean	0.92	0.83	<.001
RTB ratio, mean	0.16	0.04	<.001
Major teaching hospitals, % attending hospital with RTB ratio >0.25	21.5	5.7	<.001
Bed size, mean	595.3	429.8	<.001
High technology, %	87.0	59.1	<.001

Abbreviations: NTB, nurse-to-bed; RTB, resident-to-bed.

^a Hospital characteristics are weighted by patient. High technology is defined by the provision of coronary artery bypass graft or organ transplantation services or the presence of a burn unit.

^b P values for differences in hospital characteristics were calculated using Fisher exact test for binary covariates and Wilcoxon rank sum test for continuous ones.

ing environments between focal and control hospitals to confound the comparisons of payments) and the indirect medical expenditure adjustment (because we did not want hospitals with an increased educational burden to be penalized for teaching when comparing payments).

Statistical Analysis

Matching Algorithm

Each focal patient was treated at a hospital recognized nationally as having a good nurse working environment and an NTB ratio of 1 or greater and was matched to a control patient treated at a hospital without such recognition and with an NTB ratio less than 1. The optimal match²¹ was calculated using the ASSIGN procedure in SAS (SAS Institute).²² Our algorithm exactly matched 1 of 130 procedures inside each pair and then attempted to balance 42 patient covariates by minimizing the Mahalanobis distance^{15,23-26} between cases and control patients, including age, year of admission, sex, race, emergency admission status, transfer-in status, the propensity score, the risk score, and 31 comorbidities (eAppendix 1 in the Supplement).

Matches were performed first without viewing outcomes.^{27,28} We aimed to attain standardized differences in covariate means below 0.1. We also assessed balance using Fisher exact test for binary covariates²⁹ and Wilcoxon rank sum test for continuous ones.³⁰

Comparing Outcomes

Outcomes were compared using paired methods: for binary outcomes, McNemar test³¹; for continuous outcomes, *m*-statistics,³²⁻³⁵ including the permutational *t* test.^{32,33,36} We also used the jackknife procedure to explore the potential effect of hospital-level clustering on reported *P* values.^{37,38}

Analyzing Outcomes by Patient Risk Level

Using a data set not overlapping with our matched sample (eAppendix 2 in the Supplement), we constructed a 30-day mortality model to calculate each study patient's mortality probability. After matching, we ranked each matched pair by its average risk of mortality, forming quintiles of increasing risk, and compared outcomes between focal and control patients inside each quintile. Graphs of focal-control outcome differences by risk level were produced using LOWESS in R,³⁹ its pointwise 95% bootstrap CI,⁴⁰ and 95% joint confidence ellipse for Hotelling *T*².⁴¹⁻⁴⁴

Results

Final Patient and Hospital Sample

We identified 172 225 patients who underwent general surgery in the 3 states in 606 short-term, acute-care hospitals. The focal group had 25 752 patients in 35 hospitals recognized nationally as having both good nurse working environments and NTB ratios of 1 or greater (mean NTB ratio, 1.51). Matched controls were drawn from a cohort of 62 882 patients treated at 298 nonrecognized hospitals with NTB ratios below 1 (mean NTB ratio, 0.69).

Focal hospitals with excellent nursing environments differed from controls in many ways, as seen in Table 1. For example, 21.5% of patients (n = 5400) in the focal group attended hospitals that were major teaching hospitals with resident-to-bed ratios above 0.25 compared with 5.7% of matched control patients (n = 1420). More focal patients attended hospitals that had high-level technology available (21 823 patients [87.0%] vs 14 827 [59.1%]), and more focal patients attended large hospitals, as measured by bed size greater than 250 patients (22 286 patients [88.9%] vs 15 934 [63.5%]) (see eAppendix 7 in the Supplement for the characteristics of the hospitals where focal and control patients were treated).

Quality of the Patient Matches

Using the 25 752 general surgery patients treated in the 35 focal hospitals, we formed 25 076 pairs matched exactly for the 130 surgical procedures (97.4% of the available focal patients), with 293 of the 298 available control hospitals represented in the match. Table 2 displays some of the variables used in the match. All 130 principal procedures were matched exactly, and all other patient covariates (n = 42) were balanced, with no standardized difference after matching exceeding 0.05 SD. See eAppendix 8 in the Supplement for complete details of this extremely balanced match, including frequencies of principal procedure codes.

Outcomes

Table 3 compares outcomes of focal patients and matched control patients. Focal patients had lower 30-day mortality rates than control patients (4.8% vs 5.8%; odds ratio [OR], 0.79; 95% CI, 0.73-0.86; *P* < .001; clustered *P* value = .005) (see eAppendix 9 in the Supplement for jackknife results; see eAppendix

Table 2. Selected Matched Patient Characteristics^a

Characteristic	Focal Cases (n = 25 076)	Matched Controls (n = 25 076)	Standardized Difference After Match	P Value After Match ^b
Age, mean, y	76.3	76.3	0.00	.76
Male, No. (%)	9889 (39.4)	10 091 (40.2)	-0.02	.07
Probability of 30-d death	0.04	0.04	-0.02	.38
Propensity score for attending a focal hospital	0.32	0.32	0.04	.008
Emergency admission, No. (%)	9553 (38.1)	10 087 (40.2)	-0.04	<.001
Transfer-in, No. (%)	754 (3.0)	566 (2.3)	0.05	<.001
History, No. (%)				
Congestive heart failure	5448 (21.7)	5561 (22.6)	-0.02	.023
Myocardial infarction	2045 (8.2)	1979 (7.9)	0.01	.29
Arrhythmia	6453 (25.7)	6363 (25.4)	0.01	.36
Angina	835 (3.3)	899 (3.6)	-0.01	.12
Diabetes	6998 (27.9)	6961 (27.8)	0.00	.72
Renal failure	1461 (5.8)	1489 (5.9)	0.00	.61
COPD	5609 (22.4)	5711 (22.8)	-0.01	.28
Dementia	1604 (6.4)	1675 (6.7)	-0.01	.21

Abbreviation: COPD, chronic obstructive pulmonary disease.

^a Selected matched covariates (see full table of all 172 matched variables in eAppendix 8 in the Supplement). All patients are paired for the exact same 4-digit *International Classification of Diseases, Ninth Revision, Clinical Modification* principal procedure (n = 130), so they are not displayed here.

^b P values for differences in matched variables were calculated using Fisher exact test for binary covariates and Wilcoxon rank sum test for continuous ones.

Table 3. Patient Outcomes^a

Outcome	Focal Cases (n = 25 076)	Matched Controls (n = 25 076)	OR (95% CI) ^b	P Value	Clustered P Value
Patient outcomes (discrete), %					
30-d Mortality	4.8	5.8	0.79 (0.73 to 0.86)	<.001	.005
30-d Complications	64.4	65.3	0.95 (0.91 to 0.99)	.02	.15
30-d Failure-to-rescue ^c	7.5	8.9	0.83 (0.76 to 0.90)	<.001	<.001
Patients sent to ICU	32.9	42.9	0.55 (0.52 to 0.57)	<.001	<.001
30-d Readmission	13.5	13.4	1.01 (0.95 to 1.06)	.85	.92
30-d Readmission or death	17.9	18.7	0.94 (0.90 to 0.99)	.01	.22
Patient outcomes (continuous), m-estimate					
30-d Cost, \$	27 131	27 292	-163 (-542 to 215)	.40	.80
30-d Cost without NTB adjustment, \$ ^d	25 767	27 838	-2038 (-2413 to -1663)	<.001	<.001
30-d Medicare payment, \$	26 091	25 067	1001 (710 to 1292)	<.001	.30
30-d Medicare payment without IME and geography, \$ ^e	23 182	24 024	-851 (-1113 to -589)	<.001	.03
Length of stay, d	8.4	8.6	-0.1 (-0.3 to -0.0)	.01	.33

Abbreviations: ICU, intensive care unit; IME, indirect medical expenditure; NTB, nurse-to-bed; OR, odds ratio.

^a P values for outcomes were calculated using McNemar test for discrete outcomes and the test of the weighted m-statistic for continuous ones.

^b For continuous outcomes, the paired difference and 95% CI is reported.

^c Failure-to-rescue rates reflect the failure-to-rescue rate among focal or matched controls, while the OR and tests use only pairs in which both patients

had a qualifying failure-to-rescue event.

^d This represents the hospital's resource utilization-based costs without adding additional (or reduced) dollars for increased nursing hours per patient in hospitals with higher (or lower) NTB ratios. All costs and payments are standardized to 2008 US dollars.

^e This represents Centers for Medicare and Medicaid Services payments rendered for patient care, without IME payments or geography adjustments.

10 in the Supplement for sensitivity analysis results). Focal patients also had lower 30-day failure-to-rescue rates (7.5% vs 8.9%; OR, 0.83; 95% CI, 0.76-0.90; $P < .001$) and were in the ICU less often (32.9% vs 42.9%; OR, 0.55; 95% CI, 0.52-0.57; $P < .001$). Length of stay was slightly shorter among focal patients than matched control patients (m-estimate, 8.4 vs 8.6 days; paired difference, -0.1; 95% CI, -0.3 to -0.0; $P = .01$). Results for in-hospital outcomes were generally similar to 30-day results (eAppendix 11 in the Supplement).

Did better quality cost more? As measured by resource utilization, focal patients had similar in-hospital and 30-day

costs per patient as their controls. Thirty-day cost per patient was \$27 131 vs \$27 292 (focal vs control), a difference of -\$163 per patient pair (95% CI, -\$542 to \$215; $P = .40$). If we take away the NTB adjustment, we see a focal-control difference of -\$2038 per patient pair ($P < .001$; clustered P value $< .001$).

Payments from Medicare were higher in focal patients. Estimated 30-day payments for focal patients were \$26 091 per patient vs \$25 067, a paired difference of \$1001 per patient (95% CI, \$710 to \$1292; $P < .001$; clustered P value = .30). However, when both the geography payment adjustment and in-

Table 4. Outcome Results by Subsets of Teaching Status of the Matched Pairs^a

Hospital Type	All Patients and Hospitals					Without Major Teaching Hospitals, RTB Ratio >0.25				
	Pairs, No.	Focal Cases ^a	Matched Controls ^a	OR or Paired Difference (95% CI)	P Value ^b	Pairs, No.	Focal Cases ^a	Matched Controls ^a	OR or Paired Difference (95% CI)	P Value ^b
30-d Mortality for Focal vs Control Hospitals, %										
All hospitals	25 076	4.8	5.8	0.79 (0.73 to 0.86)	<.001	18 641	5.0	5.8	0.81 (0.74 to 0.90)	<.001
Teaching vs teaching	3823	4.6	6.4	0.67 (0.54 to 0.83)	<.001	1837	5.1	6.7	0.71 (0.52 to 0.97)	.03
Nonteaching vs nonteaching	6895	5.2	6.0	0.85 (0.72 to 1.00)	.04	6895	5.2	6.0	0.85 (0.72 to 1.00)	.04
Teaching vs nonteaching	12 313	4.5	5.6	0.76 (0.67 to 0.86)	<.001	8343	4.6	5.7	0.78 (0.67 to 0.91)	.001
Nonteaching vs teaching	2045	6.3	5.9	1.08 (0.81 to 1.44)	.62	1566	5.7	5.8	0.97 (0.69 to 1.36)	.93
ICU Use Rate for Focal vs Control Hospitals, %										
All hospitals	25 076	32.9	42.9	0.55 (0.53 to 0.57)	<.001	18 641	33.7	43.5	0.55 (0.53 to 0.58)	<.001
Teaching vs teaching	3823	32.9	45.1	0.49 (0.43 to 0.55)	<.001	1837	34.0	49.5	0.40 (0.33 to 0.47)	<.001
Nonteaching vs nonteaching	6895	37.0	41.9	0.75 (0.69 to 0.81)	<.001	6895	37.0	41.9	0.75 (0.69 to 0.81)	<.001
Teaching vs nonteaching	12 313	29.5	42.3	0.46 (0.43 to 0.49)	<.001	8343	30.0	42.4	0.47 (0.44 to 0.51)	<.001
Nonteaching vs teaching	2045	39.0	45.9	0.66 (0.56 to 0.77)	<.001	1566	38.8	50.0	0.51 (0.42 to 0.61)	<.001
Length of Stay (m-Estimate) for Focal vs Control Hospitals, d										
All hospitals	25 076	8.4	8.6	-0.1 (-0.3 to -0.0)	.01	18 641	8.4	8.5	-0.1 (-0.2 to 0.1)	.42
Teaching vs teaching	3823	8.9	9.4	-0.5 (-0.8 to -0.2)	.001	1837	8.8	9.2	-0.4 (-0.8 to 0.1)	.09
Nonteaching vs nonteaching	6895	8.2	8.3	-0.1 (-0.3 to 0.1)	.51	6895	8.2	8.3	-0.1 (-0.3 to 0.1)	.51
Teaching vs nonteaching	12 313	8.4	8.3	0.0 (-0.1 to 0.2)	.84	8343	8.4	8.3	0.1 (-0.05 to 0.3)	.13
Nonteaching vs teaching	2045	8.7	9.5	-0.7 (-1.1 to -0.3)	.001	1566	8.6	9.3	-0.6 (-1.1 to -0.2)	.008
30-d Cost/Patient (m-Estimate), \$										
All hospitals	25 076	27 131	27 292	-163 (-542 to 215)	.40	18 641	27 368	27 176	172 (-263 to 608)	.44
Teaching vs teaching	3823	28 116	28 896	-633 (-1650 to 385)	.22	1837	28 200	28 879	-625 (-2025 to 777)	.38
Nonteaching vs nonteaching	6895	27 218	26 637	531 (-160 to 1222)	.13	6895	27 218	26 637	531 (-160 to 1222)	.13
Teaching vs nonteaching	12 313	26 340	26 834	-450 (-984 to 85)	.10	8343	26 968	26 780	210 (-448 to 871)	.53
Nonteaching vs teaching	2045	29 498	29 348	57 (-1375 to 1479)	.94	1566	29 122	29 704	-658 (-2276 to 950)	.42

Abbreviations: ICU, intensive care unit; OR, odds ratio; RTB, resident-to-bed.

^a Values are expressed in the units given in each section.

^b P values are calculated using McNemar test for binary outcomes and the test of the weighted m-statistic for continuous ones.

direct medical expenditure payments were removed, 30-day payment was actually \$851 less per focal vs control patient (95% CI, -\$1113 to -\$589; $P < .001$; clustered P value = .03). In-hospital payment results were similar to the 30-day results (eAppendix 6 in the Supplement).

Analyzing Outcomes by Hospital Characteristics

The central question of this study asks whether value differences exist across hospitals selected for better or worse nursing environments and NTB ratio, but a different question is the extent to which the nurses themselves are the cause of the value differences. An explanatory variable of interest is teaching status, and all hospital characteristics associated with teaching status. We divided the 25 076 matched pairs into 4 possible combinations of teaching status of hospitals attended by

each of the 2 patients in each matched pair (focal vs control): teaching vs teaching, nonteaching vs nonteaching, teaching vs nonteaching, and nonteaching vs teaching. The resulting outcome differences can be seen in Table 4. For 30-day mortality, the focal patient advantage was maintained in all comparison combinations except when the focal patient attended a nonteaching hospital and the control attended a teaching hospital—when the odds of mortality become similar. Thirty-day costs were generally similar in the focal hospitals and controls for all 4 comparisons. Intensive care unit use was consistently less in focal vs control patients, with a very large and significant reduction in the odds of using the ICU compared with matched control patients. To verify the stability of these findings, we repeated the analysis excluding any pairs in which either patient attended a major teaching hospital (resident-

Table 5. Outcomes in Focal vs Control Nursing Environments by Patient Risk^a

Outcome (n = 25 076 Matched Pairs)	Overall Rate (n = 25 076 Pairs)	Outcomes by Patient Risk Strata				
		Lowest (n = 5015)	Lower (n = 5015)	Middle (n = 5016)	Higher (n = 5015)	Highest (n = 5015)
30-d Mortality rate, %						
Focal	4.8	0.2	0.7	1.8	4.2	17.3
Control	5.8	0.4	1.1	2.2	5.8	19.9
Difference in mortality	-1.0 ^b	-0.2	-0.3	-0.4	-1.6 ^b	-2.6 ^b
ICU use rate, %						
Focal	32.9	7.6	17.8	28.8	42.1	68.0
Control	42.9	12.5	26.4	41.2	57.2	77.0
Difference in ICU use	-10.0 ^b	-4.9 ^b	-8.6 ^b	-12.4 ^b	-15.2 ^b	-9.1 ^b
LOS, mean, d						
Focal	8.7	2.8	5.6	7.9	10.9	16.3
Control	8.8	3.1	5.8	8.2	11.2	15.8
Difference in LOS	-0.1 ^b	-0.2 ^b	-0.2 ^b	-0.3 ^c	-0.3 ^c	0.5
Comparative LOS, % difference	-1.1	-7.9	-2.6	-3.6	-2.5	3.1
30-d Cost, mean, \$						
Focal	28 119	10 919	17 900	24 561	33 513	53 701
Control	28 045	10 935	17 463	24 691	34 375	52 760
Difference in cost	74	-16	437	-130	-862	941
Comparative cost, % difference	0.3	-0.1	2.5	-0.5	-2.5	1.8
30-d Cost without NTB adjustment, mean, \$						
Focal	26 747	10 383	16 865	23 145	31 732	51 608
Control	28 585	11 191	17 899	25 263	35 061	53 511
Difference in cost	-1839 ^b	-808 ^b	-1034 ^b	-2118 ^b	-3329 ^b	-1904 ^d
Comparative cost, % difference	-6.4	-7.2	-5.8	-8.4	-9.5	-3.6

Abbreviations: ICU, intensive care unit; LOS, length of stay; NTB, nurse-to-bed. ^c P < .01.

^a P values calculated using McNemar test for binary outcomes and permutational t test for continuous outcomes.

^d P < .05.

^b P < .001.

to-bed ratio >0.25); as can be seen on the right side of Table 4, the results were similar.

Stratifying the patient pairs by hospital size or technology generally yielded results similar to those of the teaching analysis (eAppendix 12 in the Supplement).

Influence of Patient Risk

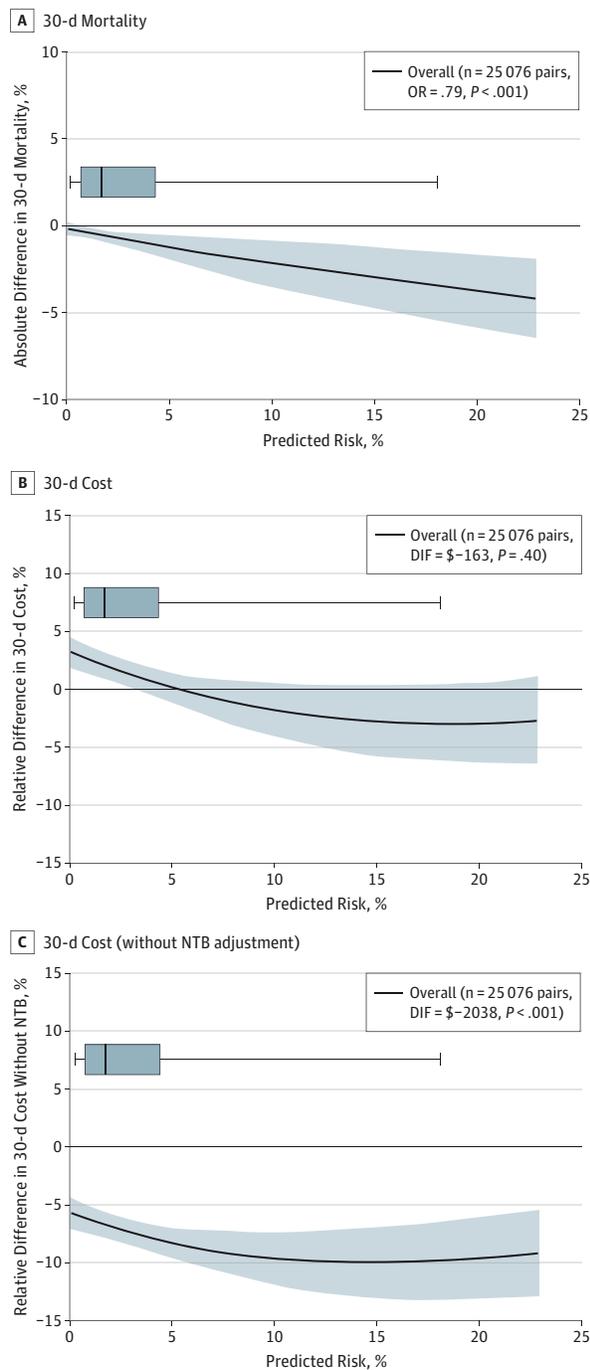
Overall, focal patients had better outcomes with costs similar to those of control patients. Do some types of patients benefit more than others? The right side of Table 5 divides the matched pairs into quintiles based on the predicted risk of 30-day mortality, that is, the risk score that was closely matched in each pair. Focal patients had lower mortality than control patients in all risk quintiles, but the difference was larger and statistically significant among higher-risk patients. In the second-highest risk quintile, mortality was 1.6% lower at focal hospitals than at control hospitals (4.2% vs 5.8%; P < .001), and in the highest-risk quintile, mortality was 2.6% lower at focal hospitals (17.3% vs 19.9%; P < .001). This trend was statistically significant (P < .001). Focal patients had 30-day cost similar to that of controls in all risk quintiles (second-highest quintile: \$33 513 vs \$34 375 [difference of -\$862]; P = .12; highest quintile: \$53 701 vs \$52 760 [difference of \$941]; P = .25), as well as similar lengths of stay. Costs without NTB adjustment were

lower at focal hospitals, and ICU use was far lower across all risk quintiles. Figure 1 displays the difference between focal and control matched pairs plotted against the initial risk of each matched pair. For 30-day mortality, focal patients are consistently below the line of equivalence that denotes a 0 difference between the focal minus control patient outcome (Figure 1A). For cost, it appears there is very little difference between groups (Figure 1B). Not adjusting for NTB differences, costs appear lower in the focal group, with savings increasing with risk (Figure 1C).

Comparing Value Across Nursing Environment by Patient Risk

In Figure 2, we compare value in the matched pairs of patients in the control (worse) and focal (better) nursing environments by patient risk. The x-axis represents the control minus focal paired difference in 30-day costs for each matched pair. The y-axis represents the control minus focal difference in 30-day mortality. The ellipses on these graphs represent the 95% joint confidence region for cost and quality mean differences.

We display 6 ellipses: 5 numbered ones including about 5015 matched sets of patients by risk quintile, and a central ellipse with a centered dot, which is based on all patients (N = 25 076) (see Appendix 13 in the Supplement for further explanation of the size of the ellipses). Ellipses crossing the

Figure 1. Thirty-Day Mortality and 30-Day Cost by Patient Risk Level

The x-axis represents the average risk of each individual matched pair; y-axis, the difference in outcome (focal-control) inside each matched pair. A point falling on the horizontal line at 0 represents no difference between outcomes of the 2 patients in the matched pair; a point falling below the line, a better outcome for the focal vs control patient. LOWESS confidence bands for the central tendency line were produced using the bootstrap method. The box plots describe the distribution of predicted risk from the fifth to the 95th percentiles. A, The mortality advantage from attending a focal hospital increases with escalating patient risk. OR indicates odds ratio. B, Only small and mostly insignificant cost differences are seen between focal and control hospitals. DIF indicates difference. C, The focal patients have lower costs when differences in the nurse-to-bed (NTB) ratio are not included in the costing formula. DIF indicates difference.

horizontal axis at 0 suggest no difference in quality. Ellipses crossing the vertical axis at 0 suggest no difference in cost. For Figure 2A, most ellipses are above the horizontal line, suggesting better quality at the focal hospitals (lower mortality than matched controls). At the same time, most ellipses also cross the vertical axis, suggesting no difference in cost. Together, there is a strong case for better value (similar cost with lower mortality) in the focal group compared with the matched controls. In Figure 2B, we see a somewhat different pattern. When we did not adjust for different NTB ratios, the second-highest risk group displayed both significantly better quality and significantly lower resource utilization.

Timing of Nursing Environment Recognition and Outcomes

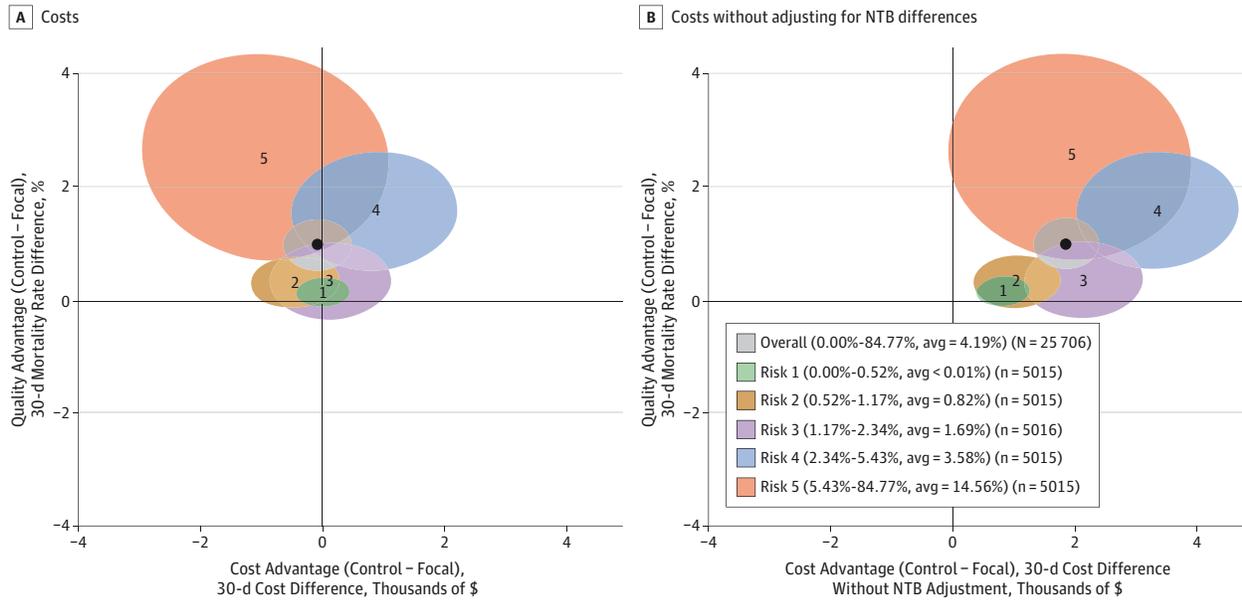
In this study, we defined a hospital with recognized excellent nursing environments if this recognition was achieved either before or including 2007 because it reflects conditions in the hospital in the recent past, as our patients were admitted between 2004 and 2006. To examine this definition more closely, we performed an analysis that excluded the subset of focal patients whose admissions occurred in hospitals that would be certified by 2007 but had not yet been certified by the year the patient was admitted. That is, for a patient to be included in the new analysis, his or her hospital had to be certified by the year the patient was admitted. Our results were unchanged. After exclusions, there were 18 212 matched pairs. The original 30-day mortality OR was 0.79 (95% CI, 0.73-0.86; P < .001; Table 3); after exclusions, the OR was 0.77 (95% CI, 0.68-0.88; P < .001). For 30-day cost, we previously found a difference of -\$163 (95% CI, -\$542 to \$215; P = .40); after exclusions, the difference was -\$138 (95% CI, -\$584 to \$307; P = .54). For ICU use, the previous OR was 0.55 (95% CI, 0.52-0.57; P < .001); after exclusions, the OR was also 0.55 (95% CI, 0.52-0.57; P < .001).

Discussion

While there is considerable evidence that a better nursing work environment is associated with better outcomes,¹⁻⁷ the question of value has remained uncertain. In this study, we asked whether 2 large groups of hospitals, defined only by different nursing environments and NTB ratios, displayed different value. We chose to examine certified hospitals with good NTB ratios because these were 2 well-known and important factors identified with better nursing environment—the accreditation of the hospital with respect to nursing environment and the most common and fundamental nurse staffing variable. Patients and referring physicians can easily observe such characteristics. When examining 30-day mortality and cost, we found that focal patients treated in better nursing environments and NTB ratios greater than or equal to 1 displayed a clear-cut advantage in value over patients treated at control hospitals. Focal patients have lower mortality with similar costs and, therefore, better value.

We also found that while all patients may benefit from hospitals that have a good nursing environment, sicker patients benefit more. Patients in the highest quintile of risk have the

Figure 2. Comparing Value Between Better (Focal) and Worse (Control) Nursing Environments by Patient Risk



The x-axis represents the difference between the control minus focal patient matched pair for 30-day costs (A) or 30-day costs without adjusting for nurse-to-bed (NTB) differences across hospitals(B). The y-axis represents the difference between control minus focal matched pairs for 30-day mortality. The ellipses on these graphs represent the 95% joint confidence region for cost and quality. For each plot, we display 6 ellipses: 5 numbered ones including about the same number of patients (n = 5015 or 5016), and a central ellipse with a centered dot that is based on all patients (N = 25 076) (see Appendix 13 in the Supplement for further explanation of the size of the ellipses). The ellipses in A

and B are identical with respect to value but differ in cost differences between focal and control patients. A, The second-highest risk group (ellipse 4) is completely above the horizontal line at y = 0, suggesting a significant advantage in quality for the focal group, while the intersection with the vertical line at x = 0 suggests that the increased costs in the focal group vs the control group did not reach statistical significance. B, This same risk group displays lower cost with better quality in the focal group compared with the matched controls. For the risk strata, avg indicates average.

largest reduction in mortality rates, but not lower costs, confirming that improved outcomes are possible for high-risk patients but expensive. Patients in the second-highest quintile of risk have substantial reduction in mortality and the largest reduction in cost, producing the highest value.

Focal hospitals also had dramatically lower rates of ICU use. This finding could be consistent with better nursing care on the floor, acting as a substitute for ICU care or other resource utilization for some patients, possibly leading to lower overall resource utilization and contributing to the business case for improving nursing environments.^{7,8}

Our analysis of value based on 30-day mortality and Medicare payments displayed results generally similar to those of the cost analyses. We observed that the 1% improvement in absolute mortality in the focal vs control population (4.8% vs 5.8%) was associated with a statistically significant CMS payment increase of about \$1000 (\$26 091 vs \$25 067), still a strong argument for excellent value.

Because our study asked whether a better nursing environment as defined by national recognition and NTB ratio could identify hospitals with better value, we purposely did not match on hospital characteristics. We found that using these 2 variables associated with the nursing environment produced 2 sets of hospitals with very different characteristics. Had we asked a different question related to whether a hospital administrator should improve a hospital's nursing envi-

ronment, as other studies have asked, then a different matching algorithm using both patient and hospital characteristics as well as a propensity score for being a recognized hospital could be used.

A limitation of our study was the use of a voluntary program of accreditation for good nursing environments as an indicator of hospital nursing work environment. Although hospitals with formal accreditation have been shown, on average, to have significantly better work environments than those without accreditation, there is known overlap in measured environments between hospitals with and without formal accreditation.^{2,11} However, our study did not use formal accreditation alone to define different nursing environments but also separated hospitals by their NTB ratio, thereby helping to reducing this overlap.

Conclusions

Patients who undergo surgery in hospitals with better nursing environments typically display lower mortality, with similar costs suggesting that better nursing environments are associated with higher value. Our results do not address whether hospitals can necessarily improve their value by improving the nursing environment; other research has investigated that question. While better outcomes and value may be owing to

other features of hospitals with good nursing, excellent nursing environments appear to provide a strong signal to patients and referring physicians for better quality, lower cost, and higher value. This is especially true for higher-risk patients, where the value of a better nursing environment appears to be greatest.

ARTICLE INFORMATION

Accepted for Publication: October 8, 2015.

Published Online: January 20, 2016.
doi:10.1001/jamasurg.2015.4908.

Author Affiliations: Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Silber); Department of Health Care Management, Wharton School, University of Pennsylvania, Philadelphia (Silber); Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia (Silber, Rosenbaum, McHugh, Even-Shoshan, Fleisher, Kelz, Aiken); Center for Health Outcomes and Policy Research, University of Pennsylvania, Philadelphia (Silber, McHugh, Smith, Aiken); Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Silber, Ludwig, Niknam, Even-Shoshan); Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Silber, Fleisher); Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia (Rosenbaum); School of Nursing, University of Pennsylvania, Philadelphia (McHugh, Aiken); Population Studies Center, University of Pennsylvania, Philadelphia (Smith, Aiken); Department of Sociology, School of Arts and Sciences, University of Pennsylvania, Philadelphia (Smith, Aiken); Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Kelz).

Author Contributions: Dr Silber had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Silber, Rosenbaum, McHugh, Smith, Even-Shoshan, Aiken.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Silber, Rosenbaum, McHugh, Niknam, Kelz, Aiken.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Silber, Rosenbaum, McHugh, Ludwig, Smith, Niknam, Aiken.

Obtained funding: Silber, Even-Shoshan.

Administrative, technical, or material support: Silber, McHugh, Niknam, Even-Shoshan, Aiken.

Study supervision: Silber, McHugh, Even-Shoshan, Aiken.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research was funded by grant R01-HS018338 from the Agency for Healthcare Research and Quality, grant R01-NR014855 from the National Institute of Nursing Research, and grant NSF SBS-10038744 from the National Science Foundation.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The findings and conclusions of this report are those of the authors and do not

necessarily represent the official position of the Agency for Healthcare Research and Quality, National Institute of Nursing Research, or the National Science Foundation.

Previous Presentation: This study was presented at the 2015 AcademyHealth Annual Research Meeting; June 13, 2015; Minneapolis, Minnesota.

Additional Contributions: Traci Frank, AA, and Alex Hill, BS, Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, provided assistance with this research; they received no compensation.

REFERENCES

- Aiken LH, Smith HL, Lake ET. Lower Medicare mortality among a set of hospitals known for good nursing care. *Med Care*. 1994;32(8):771-787.
- McHugh MD, Kelly LA, Smith HL, Wu ES, Vanak JM, Aiken LH. Lower mortality in magnet hospitals. *Med Care*. 2013;51(5):382-388.
- Lake ET, Staiger D, Horbar J, et al. Association between hospital recognition for nursing excellence and outcomes of very low-birth-weight infants. *JAMA*. 2012;307(16):1709-1716.
- Friese CR, Xia R, Ghaferi A, Birkmeyer JD, Banerjee M. Hospitals in "Magnet" Program show better patient outcomes on mortality measures compared to non-"Magnet" hospitals. *Health Aff (Millwood)*. 2015;34(6):986-992.
- Mitchell PH, Shortell SM. Adverse outcomes and variations in organization of care delivery. *Med Care*. 1997;35(11)(suppl):NS19-NS23.
- Kutney-Lee A, Stimpfel AW, Sloane DM, Cimiotti JP, Quinn LW, Aiken LH. Changes in patient and nurse outcomes associated with magnet hospital recognition. *Med Care*. 2015;53(6):550-557.
- Jayawardhana J, Welton JM, Lindrooth RC. Is there a business case for magnet hospitals? estimates of the cost and revenue implications of becoming a magnet. *Med Care*. 2014;52(5):400-406.
- Martsof GR, Auerbach D, Benevent R, et al. Examining the value of inpatient nurse staffing: an assessment of quality and patient care costs. *Med Care*. 2014;52(11):982-988.
- Rosenbaum PR. Part II: matching. In: *Design of Observational Studies*. New York, NY: Springer; 2010:153-253.
- Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41.
- Aiken LH, Havens DS, Sloane DM. The Magnet Nursing Services Recognition Program. *Am J Nurs*. 2000;100(3):26-35; quiz 35-36.
- Silber JH, Rosenbaum PR, Kelz RR, et al. Examining causes of racial disparities in general surgical mortality: hospital quality vs patient risk. *Med Care*. 2015;53(7):619-629.
- Silber JH, Rosenbaum PR, Romano PS, et al. Hospital teaching intensity, patient race, and surgical outcomes. *Arch Surg*. 2009;144(2):113-120.
- Silber JH, Romano PS, Rosen AK, Wang Y, Even-Shoshan O, Volpp KG. Failure-to-rescue: comparing definitions to measure quality of care. *Med Care*. 2007;45(10):918-925.
- Silber JH, Rosenbaum PR, Kelz RR, et al. Medical and financial risks associated with surgery in the elderly obese. *Ann Surg*. 2012;256(1):79-86.
- Silber JH, Rosenbaum PR, Ross RN, et al. Template matching for auditing hospital cost and quality. *Health Serv Res*. 2014;49(5):1446-1474.
- Silber JH, Rosenbaum PR, Ross RN, et al. A hospital-specific template for benchmarking its cost and quality. *Health Serv Res*. 2014;49(5):1475-1497.
- Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery: a study of adverse occurrence and failure to rescue. *Med Care*. 1992;30(7):615-629.
- 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(1):65-71.
- Needleman J, Buerhaus PI, Stewart M, Zelevinsky K, Mattke S. Nurse staffing in hospitals: is there a business case for quality? *Health Aff (Millwood)*. 2006;25(1):204-211.
- Rosenbaum P. Optimal matching for observational studies. *J Am Stat Assoc*. 1989;84(408):1024-1032. doi:10.1080/01621459.1989.10478868.
- SAS Institute. The ASSIGN procedure. In: *SAS/OR User's Guide: Mathematical Programming, Version 8*. Cary, NC: SAS Institute; 1999:39-54.
- Silber JH, Rosenbaum PR, Trudeau ME, et al. Multivariate matching and bias reduction in the surgical outcomes study. *Med Care*. 2001;39(10):1048-1064.
- Silber JH, Rosenbaum PR, Clark AS, et al. Characteristics associated with differences in survival among black and white women with breast cancer. *JAMA*. 2013;310(4):389-397.
- Silber JH, Rosenbaum PR, Ross RN, et al. Racial disparities in operative procedure time: the influence of obesity. *Anesthesiology*. 2013;119(1):43-51.
- Rosenbaum PR. *Design of Observational Studies*. New York, NY: Springer; 2010.
- Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26(1):20-36.
- Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2(3):808-840. doi:10.1214/08-AOAS187.
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39(1):33-38. doi:10.1080/00031305.1985.10479383.
- Hollander M, Wolfe DA. The two-sample location problem. In: *Nonparametric Statistical Methods*. 2nd ed. New York, NY: John Wiley & Sons; 1999:106-125.

31. Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press; 1975.
32. Rosenbaum PR. Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. *Biometrics*. 2007; 63(2):456-464.
33. Maritz JS. A note on exact robust confidence intervals for location. *Biometrika*. 1979;66(1):163-170. doi:10.1093/biomet/66.1.163.
34. Rosenbaum PR. Two R packages for sensitivity analysis in observational studies. *Obs Studies*. 2015; 1:1-17.
35. Huber PJ. The basic types of estimates. In: *Robust Statistics*. Hoboken, NJ: John Wiley & Sons; 1981:43-55.
36. Rosenbaum PR. Package "sensitivitymw": sensitivity analysis using weighted M-statistics. Version 1.1. R Development Core Team. <http://cran.r-project.org/web/packages/sensitivitymw/sensitivitymw.pdf>. Published July 24, 2014. Accessed May 27, 2015.
37. Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982.
38. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7(1):1-26.
39. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*. 1979;74(368):829-836. doi:10.1080/01621459.1979.10481038.
40. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986;1(1): 54-75. doi:10.1214/ss/1177013815.
41. Morrison DF. An alternative model: the Hotelling T2-test: the analysis of variance for higher-way layouts. In: *Applied Linear Statistical Methods*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall Inc; 1983:440-445.
42. Hotelling H. The generalization of student's ratio. *Ann Math Stat*. 1931;2(3):360-378.
43. Fox J, Weisberg S. *An R Companion to Applied Regression*. 2nd ed. Thousand Oaks, CA: Sage; 2011.
44. Monette G, Fox J. Ellipses, data ellipses, and confidence ellipses. <http://svitsrv25.epfl.ch/R-doc/library/car/html/Ellipses.html>. Accessed June 29, 2015.