Supplementary Online Content

Liu JB, Sosa JA, Grogan RH, et al. Variation of thyroidectomy-specific outcomes among hospitals and their association with risk adjustment and hospital performance. *JAMA Surg.* Published online November 29, 2017. doi:10.1001/jamasurg.2017.4593

eAppendix. Additional Methods

eTable 1. Patient Characteristics by Thyroidectomy-Specific Outcomes

eTable 2. Odds Ratios of Fixed Effects Included in Hierarchical Models Constructed to Profile Hospitals by Thyroidectomy-Specific Outcomes Without Thyroidectomy-Specific Variables Available

eTable 3. Model Statistics With and Without Thyroidectomy-Specific Variables Used for Risk Adjustment By Outcome

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Additional Methods

The following Common Procedural Terminology (CPT®) codes were used to identify partial, subtotal, and total thyroidectomies in this study for years 2013, 2014, and 2015: 60200, 60210, 60212, 60220, 60225, 60240, 60252, 60254, 60260, 60270, and 60271. These codes were collapsed into categories and used for case-mix adjustment (i.e., "thyroidectomy type") as follows:

Thyroidectomy type	CPT Codes
Partial	60200, 60210, 60220
Total (or subtotal), simple	60212, 60225, 60240
Completion	60260
Total (or subtotal) with neck dissection	60252, 60254
Thyroidectomy, including substernal thyroid; cervical approach	60271
Thyroidectomy, including substernal thyroid; sternal split or transthoracic approach	60270

Following standard ACS-NSQIP definitions, postoperative morbidity occurred if a patient experienced at least one of the following complications: any SSI type, wound disruption, pneumonia, unplanned intubation, prolonged ventilation >48 hours, progressive renal insufficiency, acute renal failure, urinary tract infection, cardiac arrest requiring cardiopulmonary resuscitation, acute myocardial infarction, sepsis, or septic shock. Patients who, at the time of operation, had preoperative SSI, pneumonia, sepsis, or UTI, who were intubated, or who required dialysis, were not counted as having the same occurrence postoperatively, as long as the associated condition was recorded as being present preoperatively.

Readmission, as defined in the ACS-NSQIP, is tracked as readmission to any hospital as recorded in available medical records or reported by physician, patient, caregiver or family; not only readmission to the index hospital nor only hospitals participating in the ACS-NSQIP.

eTable 1. Patient Characteristics by Thyroidectomy-Specific Outcomes

	Total (n = 14540)	Hypocalcemia (n=430)	RLN Injury (n=755)	Hematoma (n=175)
Age, mean (SD)	52.1 (15.0)	48.3 (15.2)	55.4 (14.8)	56.2 (13.8)
Female	11499 (79.1)	374 (87.0)	596 (78.9)	124 (70.9)
Race				
African American	2344 (16.1)	78 (18.1)	151 (20.0)	46 (26.3)
Asian	554 (3.8)	17 (4.0)	42 (5.6)	9 (5.1)
Caucasian	9876 (67.9)	282 (65.6)	436 (57.7)	96 (54.9)
Other	1766 (12.2)	53 (12.3)	126 (16.7)	24 (13.7)
Hispanic ethnicity	790 (5.4)	28 (6.5)	42 (5.6)	7 (4.0)
Body mass index class				
Underweight	156 (1.1)	4 (0.9)	5 (0.7)	4 (2.3)
Normal	3545 (24.4)	114 (26.5)	181 (24.0)	36 (20.6)
Overweight	4461 (30.7)	110 (25.6)	234 (31.0)	45 (25.7)
Class I obesity	3241 (22.3)	93 (21.6)	174 (23.0)	43 (24.6)
Class II obesity	1695 (11.7)	65 (15.1)	88 (11.7)	21 (12.0)
Class III obesity	1442 (9.9)	44 (10.2)	73 (9.7)	26 (14.9)
ASA physical status				
1-2	10016 (68.9)	274 (63.7)	463 (61.3)	96 (54.9)
3	4314 (29.7)	148 (34.4)	278 (36.8)	72 (41.1)
4-5	210 (1.4)	8 (1.9)	14 (1.9)	7 (4.0)
Not directly admitted from home	87 (0.6)	9 (2.1)	7 (0.9)	2 (1.1)
Ventilator dependence	28 (0.2)	4 (0.9)	0 (0.0)	2 (1.1)
COPD	330 (2.3)	10 (2.3)	18 (2.4)	6 (3.4)
Ascites	8 (0.1)	2 (0.5)	1 (0.1)	0 (0.0)
CHF	55 (0.4)	4 (0.9)	2 (0.3)	2 (1.1)
Renal failure	72 (0.5)	10 (2.3)	5 (0.7)	4 (2.3)
Disseminated cancer	120 (0.8)	6 (1.4)	16 (2.1)	3 (1.7)
Chronic steroids	380 (2.6)	16 (3.7)	25 (3.3)	8 (4.6)
Weight loss	77 (0.5)	4 (0.9)	6 (0.8)	2 (1.1)
Bleeding disorder	158 (1.1)	2 (0.5)	7 (0.9)	7 (4.0)
Smoking	2122 (14.6)	80 (18.6)	120 (15.9)	25 (14.3)
Hypertension	5566 (38.3)	153 (35.6)	336 (44.5)	77 (44.0)
Diabetes				
Oral	1300 (8.9)	28 (6.5)	69 (9.1)	17 (9.7)
Insulin	645 (4.4)	16 (3.7)	39 (5.2)	10 (5.7)
Dyspnea	875 (6.0)	39 (9.1)	60 (7.9)	20 (11.4)
Dependent functional status	82 (0.6)	5 (1.2)	4 (0.5)	3 (1.7)
Preoperative SIRS/sepsis	60 (0.4)	9 (2.1)	6 (0.8)	2 (1.1)
Emergency	32 (0.2)	2 (0.5)	1 (0.1)	0 (0.0)
Outpatient	8487 (58.4)	182 (42.3)	390 (51.7)	67 (38.3)
Indication				

Single nodule	5062 (34.8)	84 (19.5)	194 (25.7)	58 (33.1)
Multinodular goiter	6196 (42.6)	186 (43.3)	327 (43.3)	71 (40.6)
Graves' disease	906 (6.2)	58 (13.5)	50 (6.6)	12 (6.9)
Differentiated malignancy	1761 (12.1)	83 (19.3)	150 (19.9)	23 (13.1)
Poorly differentiated or other malignancy	167 (1.2)	7 (1.6)	7 (0.9)	4 (2.3)
Other	448 (3.1)	12 (2.8)	27 (3.6)	7 (4.0)
Clinical toxicity	1138 (7.8)	54 (12.6)	74 (9.8)	21 (12.0)
Previous neck surgery	1575 (10.8)	44 (10.2)	106 (14.0)	25 (14.3)
Minimally-invasive approach	117 (0.8)	2 (0.5)	9 (1.2)	0 (0.0)
Central neck dissection performed	3789 (26.1)	161 (37.4)	189 (25.0)	47 (26.9)
Neoplasm	8242 (56.7)	272 (63.3)	472 (62.5)	94 (53.7)
Pathologic T classification				
N/A	8408 (57.8)	196 (45.6)	383 (50.7)	106 (60.6)
x	252 (1.7)	10 (2.3)	9 (1.2)	0 (0.0)
0	207 (1.4)	5 (1.2)	11 (1.5)	3 (1.7)
1	3403 (23.4)	125 (29.1)	188 (24.9)	36 (20.6)
2	883 (6.1)	25 (5.8)	45 (6.0)	4 (2.3)
3	1262 (8.7)	60 (14.0)	94 (12.5)	23 (13.1)
4	125 (0.9)	9 (2.1)	25 (3.3)	3 (1.7)
Multifocal				
Unilateral	1043 (7.2)	35 (8.1)	60 (7.9)	13 (7.4)
Bilateral	1324 (9.1)	74 (17.2)	95 (12.6)	18 (10.3)
Pathologic N classification				
N/A	8710 (59.9)	199 (46.3)	395 (52.3)	107 (61.1)
x	2474 (17.0)	80 (18.6)	137 (18.1)	28 (16.0)
0	2110 (14.5)	73 (17.0)	114 (15.1)	19 (10.9)
1	1246 (8.6)	78 (18.1)	109 (14.4)	21 (12.0)
Pathologic M classification				
N/A	10937 (75.2)	274 (63.7)	504 (66.8)	126 (72.0)
1	61 (0.4)	2 (0.5)	8 (1.1)	2 (1.1)
Thyroidectomy				
Partial	5154 (35.5)	28 (6.5)	190 (25.2)	52 (29.7)
Total (or subtotal), simple	6151 (42.3)	269 (62.6)	338 (44.8)	82 (46.9)
Completion	709 (4.9)	18 (4.2)	51 (6.8)	13 (7.4)
Total (or subtotal), neck dissection	1872 (12.9)	95 (22.1)	121 (16.0)	24 (13.7)
Substernal goiter, cervical	587 (4.0)	17 (4.0)	50 (6.6)	3 (1.7)
Substernal goiter, thoracic	67 (0.5)	3 (0.7)	5 (0.7)	1 (0.6)
Parathyroid autotransplantation	1316 (9.1)	59 (13.7)	96 (12.7)	15 (8.6)

RLN: recurrent laryngeal nerve; ASA: American Society of Anesthesiologists; COPD: chronic obstructive pulmonary disease; CHF; congestive heart failure; SIRS: systemic inflammatory response syndrome; TNM refers to AJCC 7th ed. tumor classification.
All values are expressed as N (%) unless otherwise noted.

eTable 2. Odds Ratios of Fixed Effects Included in Hierarchical Models Constructed to Profile Hospitals by Thyroidectomy-Specific Outcomes Without Thyroidectomy-Specific Variables Available

Hypocalcemia			RLN Injury			Hematoma		
Step*	Variable	OR (95% CI)	Step*	Variable	OR (95% CI)	Step*	Variable	OR (95% CI)
1	Thyroidectomy type		1	Race		1	Outpatient	0.44 (0.31-0.63)
	Partial	0.18 (0.12-0.28)		Caucasian	Ref	2	ASA class	
	Total (or subtotal), simple	Ref		Asian	1.63 (1.14-2.33)		1-2	Ref
	Completion	0.49 (0.29-0.83)		African American	1.40 (1.13-1.74)		3	1.33 (0.95-1.87)
	Total (or subtotal), neck dissection	0.77 (0.48-1.22)		Other/Unknown	1.21 (0.88-1.68)		4+	1.96 (0.84-4.55)
	Substernal goiter, cervical	0.60 (0.35-1.03)	2	Age	1.02 (1.01-1.02)	3	Race	
	Substernal goiter, thoracic	0.37 (0.10-1.39)	3	Work RVUs	1.05 (1.01-1.10)		Caucasian	Ref
2	Alkaline phosphatase >125 IU/L	2.54 (1.72-3.76)	4	Disseminated cancer	2.41 (1.37-4.25)		Asian	1.82 (0.90-3.69)
3	Age	0.98 (0.98-0.99)	5	Hispanic ethnicity			African American	2.04 (1.40-2.95)
4	Renal failure	5.38 (2.52-11.47)		No	Ref		Other/Unknown	1.34 (0.82-2.18)
5	SIRS/Sepsis	3.81 (1.61-9.03)		Yes	1.20 (0.83-1.73)	4	Hematocrit >45	2.02 (1.28-3.20)
6	Outpatient	0.44 (0.35-0.57)		Unknown	0.85 (0.60-1.21)	5	Age	1.01 (1.00-1.02)
7	Female	1.85 (1.37-2.50)	6	Thyroidectomy type		6	Bleeding disorder†	2.48 (1.10-5.62)
8	Not admitted from home	2.82 (1.20-6.62)		Partial	Ref	7	Wound classification	
9	Dyspnea	1.41 (0.97-2.04)		Total (or subtotal), simple	1.37 (1.08-1.75)		Clean or clean contaminated	Ref
10	Work RVU	1.08 (1.03-1.14)		Completion	1.35 (0.88-2.08)		Contaminated or dirty	4.43 (1.01-19.46)
				Total (or subtotal), neck dissection	1.16 (0.69-1.96)			
				Substernal goiter, cervical	1.76 (1.15-2.68)			
				Substernal goiter, thoracic	1.10 (0.38-3.24)			

RLN: recurrent laryngeal nerve; SIRS: systemic inflammatory response syndrome; ASA: American Society of Anesthesiologists

* Steps represent point when variable is selected in a logistic, forward stepwise regression process (entrance criteria, $p<0.05$). Only standard ACS NSQIP variables were available for selection. Selected variables are subsequently included as fixed effects for risk adjustment when evaluating hospital performance for each outcome studied.

† A preoperative bleeding disorder is present in patients with an underlying hematologic disorder (e.g., hemophilia) or on chronic anticoagulation (e.g., warfarin).

eTable 3. Model Statistics With and Without Thyroidectomy-Specific Variables Used for Risk Adjustment By Outcome

Outcome	C-statistic		Hosmer-Lemeshow test statistic (p value)		Brier score	
	With	Without	With	Without	With	Without
Hypocalcemia	0.817	0.814	6.92 (0.54)	5.92 (0.66)	0.029	0.029
Recurrent Laryngeal Nerve Injury	0.772	0.765	13.26 (0.10)	18.69 (0.02)	0.050	0.050
Hematoma	0.704	0.704	8.97 (0.34)	8.97 (0.34)	0.013	0.013
Morbidity	0.811	0.804	8.82 (0.36)	12.71 (0.12)	0.013	0.013
Surgical Site Infection	0.840	0.836	12.34 (0.14)	12.74 (0.12)	0.004	0.004
Readmission	0.706	0.683	18.72 (0.02)	10.59 (0.23)	0.024	0.024

The C-statistic is used to assess discrimination, which represents the probability that an individual who experienced the event had a higher model-assigned risk than an individual who did not. The C-statistic ranges from 0.5 (discrimination equal to chance) to 1.0 (perfect discrimination).

The Hosmer-Lemeshow test assesses calibration, which refers to the agreement between observed and predicted outcomes and the consistency of that agreement across the range of predicted risk. A non-significant test reflects adequate calibration. It utilizes a χ^2-test statistic and is thus dependent on sample size – larger samples sizes will result in statistical significance.

The Brier score assesses model accuracy, or discrimination and calibration simultaneously. The Brier score is the average squared difference between the patient's predicted probability for the event and the observed event (0 or 1, for non-event and event, respectively). Values closer to 0 reflect improved model specification.