ANNALS OF SURGERY
Vol. 228, No. 3, 320–330

The Importance of Surgeon Experience for Clinical and Economic Outcomes From Thyroidectomy

Julie Ann Sosa, MD,*† Helen M. Bowman, MS,‡ James M. Tielsch, PhD,§ Neil R. Powe, MD, MPH, MBA,†‖¶#
Toby A. Gordon, ScD,*‡ and Robert Udelsman, MD*

From the Departments of Surgery and Medicine‖ and the Robert Wood Johnson Clinical Scholars Program, The Johns Hopkins University School of Medicine†; The Johns Hopkins Hospital‡; and the Departments of International Health,§ Epidemiology,¶ and Health Policy and Management,# The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland*

Objective
To determine whether individual surgeon experience is associated with improved short-term clinical and economic outcomes for patients with benign and malignant thyroid disease who underwent thyroid procedures in Maryland between 1991 and 1996.

Summary Background Data
There is a prevailing belief that surgeon experience affects patient outcomes in endocrine surgery, but there is a paucity of objective evidence outside of clinical series published by experienced surgeons that supports this view.

Methods
A cross-sectional analysis of all patients who underwent thyroidectomy in Maryland between 1991 and 1996 was conducted using a computerized statewide hospital discharge data base. Surgeons were categorized by volume of thyroidectomies over the 6-year study period: A (1 to 9 cases), B (10 to 29 cases), C (30 to 100 cases), and D (>100 cases). Multivariate regression was used to assess the relation between surgeon caseload and in-hospital complications, length of stay, and total hospital charges, adjusting for case mix and hospital volume.

Results
The highest-volume surgeons (group D) performed the greatest proportion of total thyroidectomies among the 5860 discharges, and they were more likely to operate on patients with cancer. After adjusting for case mix and hospital volume, highest-volume surgeons had the shortest length of stay (1.4 days *vs.* 1.7 days for groups B and C and 1.9 days for group A) and the lowest complication rate (5.1% *vs.* 6.1% for groups B and C and 8.6% for group A). Length of stay and complications were more determined by surgeon experience than hospital volume, which had no consistent association with outcomes.

Conclusions
Individual surgeon experience is significantly associated with complication rates and length of stay for thyroidectomy.

Disorders of the thyroid gland occur in 3% to 5% of the population and are the second most prevalent endocrine diseases, behind diabetes mellitus.[1] Although goiters have been described since antiquity, until the late 1800s surgical treatment was reserved for patients with impending death from suffocation because of fears of devastating surgical complications, including hemorrhage and infection. Until 1850, 70 thyroidectomies are known to have been performed, with a mortality rate of 41%.[2]

With the introduction of anesthesia, antisepsis, and fine surgical instruments, and the pioneering work of Kocher, Halsted, Lahey, and Crile, thyroid surgery became a safer alternative, offering the chance of cure to many patients.[3–6] Today thyroidectomy is a routine procedure performed for

Presented at the 118th Annual Meeting of the American Surgical Association, Palm Beach, Florida, April 1998.
Supported by the Robert Wood Johnson Foundation.
Address reprint requests to Robert Udelsman, MD, Chief, Division of Endocrine and Oncologic Surgery, Blalock 688, The Johns Hopkins Hospital, 600 N. Wolfe St., Baltimore, MD 21287.
Accepted for publication April 1998.

thyroid cancer, multinodular goiter, and select patients with hyperthyroidism.[7] Overall, death from thyroid surgery is rare, and the incidence of recurrent laryngeal nerve injury, neck hematoma, thyroid storm, and hypoparathyroidism is low, although sequelae such as vocal cord paralysis and airway obstruction can be devastating.[8,9] These reported complication rates vary dramatically, ranging from 0% to 14% for permanent recurrent laryngeal nerve injury and from 1.2% to 11% for permanent hypoparathyroidism.[10] It has been suggested that these rates are higher after more extensive resections (*e.g.,* total thyroidectomy) and redo procedures, and when thyroid surgery is performed by less-experienced surgeons.[11]

Using administrative data, researchers have established an association between volume and clinical or economic outcomes for a variety of procedures. Improved outcomes may be the result of the improved technique that results from greater experience, selective referral of patients to high-volume centers or surgeons, or treatment of sicker patients by low-volume providers. Sometimes it is increased hospital volume that is associated with improved patient outcomes, as in the case of pancreaticoduodenectomy, cholecystectomy, coronary arteriography, and angioplasty, and sometimes it is surgeon volume that is significant, as for arthroplasty, coronary bypass surgery, repair of abdominal aortic aneurysms, partial gastrectomies, and colectomies.[6,12–17]

There is a paucity of objective evidence outside of clinical series published by endocrine surgeons to support a consistent association between surgeon experience and patient outcomes. There is no benchmark for comparison that represents the outcomes of thyroid patients operated on by community surgeons. In addition, the evidence is contradictory; some small series suggest that low-volume surgeons, well-supervised surgical trainees, and surgeons at community hospitals can obtain excellent clinical outcomes.[4,18–21]

We conducted a statewide study of surgical procedures for benign and malignant thyroid disease to measure the effect of individual surgeon volume on clinical and economic outcomes, including in-hospital complications, length of stay, and hospital charges. Because several factors may confound these assessments, consideration was given to controlling for additional patient and hospital characteristics in the analysis.

METHODS

Hospital and Patient Characteristics

The study design was a cross-sectional analysis of hospital discharge data from nonfederal acute care hospitals in Maryland collected by the Maryland Health Services Cost Review Commission (HSCRC). These publicly available data were used to identify all adult patients (18 years of age and older) who underwent a primary procedure that was a thyroidectomy for benign or malignant thyroid disease in Maryland between January 1991 and December 1996. Thyroid procedures included in the analysis included unilateral thyroid lobectomy (International Classification of Diseases [ICD] procedure code 06.2), other partial thyroidectomy (ICD-9 code 06.3), complete thyroidectomy (ICD-9 code 06.4), substernal thyroidectomy (ICD-9 code 06.5), excision of lingual thyroid (ICD-9 code 06.6), and other operations on thyroid glands (ICD-9 code 06.98).

The main independent variable in this study, individual surgeon volume, was modeled as a categorical variable. Surgeons were included in the analysis if they performed at least one thyroidectomy during the study period. Surgeons were categorized according to their total volume of thyroidectomies during the 6-year study period as follows: surgeons in group A performed 1 to 9 thyroidectomies, group B 10 to 29, group C 30 to 100, and group D more than 100.

Among the diagnoses considered in the analysis were thyroid adenoma (ICD-9 code 226, benign neoplasm–thyroid), other benign disease (ICD-9 code 241, nontoxic nodular goiter; 242, thyrotoxicosis; 245, thyroiditis; 240, goiter simple–not otherwise specified; 246, other disorders of thyroid), and malignancy (ICD-9 code 193, malignant neoplasm of thyroid). Other independent variables were patient age, gender, race (white, black, other), payer status (Medicare, Medicaid, commercial insurance, HMO, other/unknown), time period (1991 to 1993, 1994 to 1996), and comorbidity score as determined using the Dartmouth–Manitoba adaptation of the Charlson comorbidity index.[22,23] The comorbidity score was treated as a continuous variable in our analysis. Also considered in the models were hospital volume (1 to 99, 100 to 199, 200 to 300, >300 thyroidectomies per year), urgency of admission (urgent/emergent, elective, other/unknown), and the place of patient residence (Baltimore inner city, eastern Maryland, central Maryland, southern Maryland, western Maryland, Washington DC suburbs, all other states, foreign countries, and other/unknown).

Patient Outcomes

The outcomes of interest were in-hospital complications, mean length of stay, and mean total hospital charges. In-hospital death was not considered because it was extremely rare (only three in-hospital deaths in our study population). Complications were derived from patients' secondary diagnoses and procedures and included unanticipated conditions believed to be related to the index hospitalization (described as "surgical complications," "abnormal reactions," or "acute" events in the ICD-9 coding system; see Appendix A) to exclude preexisting comorbid conditions. We included complications that were directly related to surgery (*e.g.*, recurrent laryngeal nerve injury) as well as those that were indirectly related (*e.g.*, allergic drug reaction). Therefore, complication rates reported in this study may be higher than in those studies that report only complications directly related to surgery. It was not possible to measure compli-

cation severity, so the outcome was treated as a dichotomous variable (no complications *vs.* one or more).

Complications included diagnoses and procedures in 10 broad categories: endocrine (hypoparathyroidism), wound (infection/cellulitis, hematoma requiring exploration, debridement, or evacuation), respiratory (pneumothorax, upper respiratory tract infection, pharyngitis, mechanical ventilation >96 hours), urologic (infection), hematologic/vascular (hemorrhage, blood transfusion), cardiovascular (cardiac arrest), neurologic (recurrent laryngeal nerve injury), drug reaction/allergy, gastrointestinal, and other (abnormal reactions). Four hundred fifty-seven of 561 medical records (82%) were reviewed at the highest-volume hospital to determine if complications identified in the discharge abstract registry were accurately described compared to the hospital chart. A review of the number of complications revealed a positive predictive value of 82% and a negative predictive value of 98%.

Hospital charges were adjusted for inflation based on the appropriate annual Health Care Financing Administration input price indices, and results are presented in constant 1996 dollars.[24] Because hospital charges are strictly regulated in Maryland, they serve as a reasonable proxy for actual costs. The average cost-to-charge ratio in Maryland hospitals is about 0.75.[25]

Statistical Analyses

The distributions of patient characteristics among provider groups were compared using analysis of variance for the continuous variables (age and comorbidity score) and the χ^2 statistic for categorical variables. Bivariate analyses were used to determine which variables were associated with our outcomes of interest. These analyses guided the selection of variables for adjustment in the multivariate regression models.

Multiple linear regression was used to assess how length of stay and mean total hospital charges differed between surgeon volume groups after adjusting for potentially confounding variables that included patient age, race, comorbidity score, thyroid diagnosis and procedure, insurance status, hospital volume, and time period. Because the distribution of length of stay and hospital charges was skewed to the right, we performed a natural log transformation to achieve a more normal distribution. To estimate adjusted length of stay and total charges for each group, we transformed the data back to their original scales by exponentiating values predicted by the models. Multiple logistic regression was used to model the risk of in-hospital complications between surgeon volume groups, adjusting for the same patient case mix and hospital volume variables indicated above. For all three outcomes (complications, length of stay, and charges), the full data set was used to calculate regression coefficients for the independent variables. These coefficients were then applied to patients within the different surgeon volume groups to derive adjusted outcomes. Finally, patients were stratified by thyroid diagnosis and procedure, and separate multivariate subgroup analyses were performed to measure the relation between surgeon volume and outcomes in these subgroups.

All statistical inferences pertaining to mean length of stay and hospital charges are based on the log-transformed data. Probability values greater than 0.05 are reported as nonsignificant. All probability values are the results of two-sided tests. Data management and analysis were performed using Paradox 4.5 (Borland International, Scotts Valley, CA) and STATA 5.0 (STATA release 5, College Station, TX), respectively.

RESULTS

Patient Characteristics

A total of 5860 patients in Maryland underwent thyroid procedures between 1991 and 1996 in 52 nonfederal acute care hospitals. Their average age was 48.6 years, and the majority were women (80.5%), white (72.5%), and otherwise healthy (less than one comorbid condition, on average) (Table 1). Because the study population represents the subset of all thyroid patients who underwent surgery, thyroid carcinoma was more prevalent in our study than in the population of patients with any thyroid disease; one quarter had a thyroid malignancy.

Surgeon Characteristics

The experience of 658 surgeons was analyzed. There were 517 surgeons (1457 cases) in group A, 98 (1906 cases) in B, 39 (1651 cases) in C, and 4 (846 cases) in D. Surgeons included in this study (one or more thyroidectomies between 1991 and 1996) performed a median of 25 thyroidectomies, but nearly two thirds of the surgeons performed fewer than one thyroidectomies, on average, per year. The highest-volume surgeons did a disproportionately large proportion of all thyroidectomy (Fig. 1): although group D surgeons represented less than 1% of surgeons, they did 14.4% of the cases. In contrast, group A surgeons represented more than 75% of surgeons but did only 25% of the cases. Over time, the highest-volume surgeons appear to have increased their regional referral base: their share of thyroidectomies rose from 11.9% in 1991 to 1993 to 17.6% in 1994 to 1996. At the same time, the lowest-volume surgeons lost share, from 27.3% to 20%, and medium-volume surgeons' share remained relatively stable.

Overall, the highest-volume surgeon group (D) performed the most complex cases; these surgeons did the greatest proportion of total thyroidectomies (29% in group D *vs.* 15% in group A), and they were more likely to operate on patients with cancer (31% in group D *vs.* 23% in group A). The highest-volume surgeons also had significantly younger patients and a slightly smaller proportion of female

Table 1. PATIENT CHARACTERISTICS BY SURGEON VOLUME GROUPS

Patient Characteristics	State Average	Surgeon Volume Groups A 1–9 cases n = 1457	B 10–29 cases n = 1906	C 30–100 cases n = 1651	D > 100 cases n = 846	p value*
Age (mean years)	48.6	49.9	49.0	48.1	46.2	<0.001
Gender: Female (%)	80.5	79.7	81.3	81.5	78.3	NS
Race (%)†						<0.001
White	72.5	66.7	74.1	75.0	73.4	
Black	23.1	28.7	22.1	20.1	21.9	
Other	4.4	4.7	3.8	4.9	4.7	
Comorbidities (mean score)	0.41	0.47	0.36	0.38	0.43	NS
Case distribution by period (%)‡						<0.001
1991–93		27.3	32.9	28.0	11.9	
1994–96		20.0	33.2	29.2	17.6	
Diagnosis (%)†						<0.001
Adenoma	23.6	24.1	27.3	23.5	14.4	
Hyperplasia (other benign)	51.4	52.6	50.0	50.0	55.1	
Cancer	25.1	23.3	22.8	26.5	30.5	
Procedure (%)†§						<0.001
Unilateral lobectomy	46.2	46.1	49.6	45.9	39.1	
Other partial thyroidectomy	30.1	32.2	30.4	28.7	28.8	
Total thyroidectomy	19.5	15.1	15.8	23.0	28.8	
Substernal thyroidectomy	3.8	5.7	3.7	2.5	3.0	
Other procedures NOS	0.4	0.9	0.5	0.1	0.2	
Payment source (%)†						<0.001
Commercial insurance	74.1	67.0	72.9	78.5	79.0	
Medicare	17.5	20.9	18.7	16.7	11.9	
Medicaid	5.0	8.2	4.4	3.3	4.4	
Other‖	3.3	3.9	4.0	1.5	4.7	
Place of residence (%)						<0.001
Out-of-state	7.4	4.8	8.1	7.4	10.3	

Key: NS = not significant; NOS = not otherwise specified.
* Based on X^2 analysis.
† Columns might not add to 100% due to rounding.
‡ Rows might not add to 100% due to rounding.
§ Radical neck dissection (ICD-9 codes 40.40–40.42) performed with lobectomy (1 case), other partial- (16 cases), total- (84 cases), and substernal thyroidectomy (3 cases).
‖ Self-pay, other government, workers' compensation, no charge.

patients, which probably reflect the higher prevalence of cancer in group D cases.

Patient Outcomes

Overall, there was a significant pattern of association between increasing surgeon volume and improved outcomes (Table 2). The highest-volume surgeons had the fewest complications and the shortest length of stay both before and after adjustment for differences in patient case mix (age, race, comorbidity score, insurance status, thyroid diagnosis and procedure), hospital volume, and time period. They were minimally more expensive than moderate-volume surgeons (groups B and C), and their charges were similar to lowest-volume surgeons (group A).

The in-hospital complication rate after thyroidectomy in Maryland was 7.4%. Of 5860 patients, 431 experienced

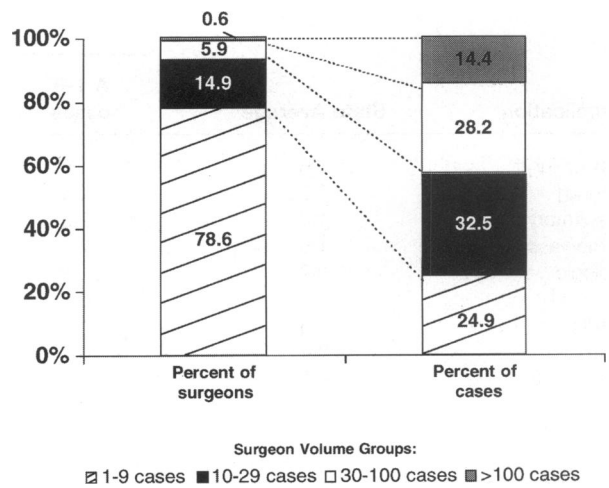

Fig. 1. Summary of the distribution of thyroid surgeons and cases by the four surgeon volume groups.

Table 2. UNADJUSTED AND ADJUSTED CLINICAL AND ECONOMIC OUTCOMES FROM THYROIDECTOMY BY SURGEON VOLUME GROUP

Outcomes	Surgeon Volume Groups			
	A 1–9 cases	B 10–29 cases	C 30–100 cases	D > 100 cases
Complication rate				
Unadjusted (%)	10.1†	6.7†	6.9†	5.9
Adjusted (%)*	8.6†	6.1†	6.1†	5.1
Length of stay				
Unadjusted (days)	2.8‡	2.1‡	2.2‡	1.7
Adjusted (days)*	1.9‡	1.7‡	1.7‡	1.4
Hospital charges				
Unadjusted ($)	$5078†	$4084†	$4016†	$4777
Adjusted ($)	$3901	$3693†	$3585†	$3950

* Adjusted for patient age, race, comorbidities, insurance status, diagnosis, procedure, surgeon and hospital volume.
† $p < 0.001$ compared to Group D (ref).
‡ $p < 0.05$ compared to Group D (ref).
Ref = Reference category when dummy variables used to calculate P values.

postsurgical complications; 339 had one complication, 77 had two, 10 had three, and 5 had four. Differences in unadjusted postsurgical morbidity rates across surgeon volume groups were statistically significant for laryngeal nerve injury, wound, and respiratory complications (Table 3). Of the 46 cases of vocal cord paralysis, 29 were unilateral; of these, 19 were partial and 10 were total. Overall, rates of postsurgical hypoparathyroidism appeared low and were not significantly different across volume groups. This could be because most cases of hypoparathyroidism are diagnosed after discharge and therefore are not captured in this data set. Recurrent laryngeal nerve injury was most common in thyroid cancer patients (1.4%), after substernal thyroidectomy (4.6%), and total thyroidectomy (1.1%).

In addition to surgeon caseload, patient age, diagnosis, and procedure were associated with all three patient outcomes (Table 4). After adjustment, thyroid patients 70 years of age and older had significantly higher complication rates, longer length of stay, and higher hospital charges than did similar, younger patients. Rates of all 10 categories of complications were elevated, probably because of the greater severity of illness or complexity of comorbidities that is typical among the elderly. Patients with thyroid cancer and patients who underwent total or substernal thyroidectomy also had worse outcomes than similar patients with benign diagnoses who underwent less extensive operations. Higher complication rates were associated with longer hospital stay, but it was not possible to determine from this cross-sectional analysis whether this was a causal association.

Subgroup Analyses

Stratified multivariate analyses demonstrated that increased surgeon volume was an especially important predictor of outcomes in the management of patients who had more severe diagnoses, such as thyroid cancer (Table 5), and who required more extensive and technically challenging operations, such as total thyroidectomy (Table 6). Although there were not significant differences in complication rates and charges between surgeon volume groups in

Table 3. COMPLICATION RATES (%) AFTER THYROIDECTOMY BY SURGEON VOLUME GROUP

Complication*	State Average	Surgeon Volume Groups			
		A 1–9 cases	B 10–29 cases	C 30–100 cases	D > 100 cases
Nerve injury†	0.8	1.5	0.5	0.8	0.4
Wound†	0.8	1.4	0.5	0.7	0.7
Respiratory†	1.5	2.3	1.4	1.2	1.0
Heme/vascular	1.5	2.0	1.6	1.2	1.4
Urologic	0.9	1.3	0.7	0.7	1.2
GI	0.8	0.7	0.8	0.9	0.6
Cardiac	0.5	0.7	0.3	0.7	0.5
Drug	0.4	0.7	0.3	0.3	0.8
Hypoparathyroidism	0.3	0.2	0.4	0.4	0.2
Other	1.7	2.3	1.5	1.8	1.3
Overall†	7.4	10.1	6.7	6.9	5.9

* For description of categories of complications, see Methods section.
† $p < 0.05$ based on X^2 analysis.

Table 4. OTHER FACTORS ASSOCIATED WITH OUTCOMES FROM THYROIDECTOMY

Factor	Adjusted Outcomes		
	Complication Rate (%)	Length of Stay (Days)	Charges ($)
Patient age			
< 70 years	6.3 ref	1.7 ref	3697 ref
≥ 70 years	10.3*	2.0†	4195†
Diagnosis			
Adenoma	4.8 ref	1.6 ref	3496 ref
Other benign	6.6‡	1.7†	3733†
Cancer	8.1†	1.9†	4029†
Procedure			
Lobectomy	5.8 ref	1.6 ref	3503 ref
Other subtotal	6.0*	1.7†	3597‡
Total thyroidectomy	8.4*	2.0†	4534†
Substernal thyroidectomy	10.8*	2.2†	4461†

Outcomes adjusted for surgeon volume, patient age, race, diagnosis, and procedure, hospital volume, and time period.
Key: NS = not significant; ref = reference category when dummy variables used to calculate p values in multivariate regression model.
* $p < 0.01$ compared to Group D (ref).
† $p < 0.001$ compared to Group D (Ref).
‡ $p < 0.05$ compared to Group D (Ref).
Ref = reference category when dummy variables used to calculate p values in multivariate regression model.

the management of adenomas, highest-volume surgeons (group D) had one-third fewer complications than their lowest-volume colleagues (group A) when they operated for other benign conditions and nearly two-thirds fewer complications than group A for cancer. Hospital charges were similar across volume groups for other benign conditions, but group D's charges for the management of thyroid cancer were lowest by a significant margin. Highest-volume surgeons had the shortest length of stay regardless of diagnosis, and all differences were statistically significant.

This pattern was borne out in Table 6, where the most striking differences in patient outcomes by surgeon volume group were seen for the most complex procedures. For total thyroidectomy, group D had nearly 75% fewer postsurgical complications ($p < 0.001$) and the lowest charges ($p < 0.05$ *vs.* group B, $p < 0.001$ *vs.* group A).

Thyroidectomy complications and length of stay were more determined by surgeon than hospital volume. Hospital volume had no consistent association with outcomes. This was tested in a subgroup analysis of the 2094 (36%) surgeons with operating privileges at more than one hospital, which revealed that their patients' length of stay, for example, was associated with the volume of the surgeon's thyroid practice rather than the number of procedures done at the hospital where the operation was performed. As surgeon volume increased from group A to group D among surgeons operating at multiple hospitals, length of stay decreased, from 2.5 to 2.1 to 2.1 to 1.7 days. However, as hospital volume increased, there was no consistent pattern in length of stay, which went from 2.2 to 2.2 to 2 back to 2.3 days.

The highest-volume surgeon in Maryland performed 346 thyroidectomies over the 5-year study period; the other high-volume surgeons in group D did 199, 171, and 130 procedures. The highest-volume surgeon had the lowest complication rate (2.4%; $p < 0.01$), adjusted length of stay (1.1 days; $p < 0.001$), and hospital charges ($2990 *vs.* $3620 to $4420; $p < 0.001$) compared to all surgeon volume groups.

DISCUSSION

In this analysis of statewide discharge abstract data for patients who underwent thyroidectomy in Maryland between 1991 and 1996, individual surgeon experience, rather than hospital experience, was significantly associated with complication rates and length of stay. This relation was observed in disease and procedure subgroups, and it remained significant after adjustment for patient case mix and time period. The association between surgeon procedure volume and outcomes has important implications for en-

Table 5. ADJUSTED OUTCOMES OF SURGEON VOLUME GROUPS BY DIAGNOSIS

Outcome, by Surgeon Volume Group (Cases/yr)	Diagnosis		
	Adenoma (n = 1381)	Other Benign (n = 3009)	Cancer (n = 1470)
Complications	(%)	(%)	(%)
A (1–9 cases)	5.7 NS	9.1*	12.9‡
B (10–29 cases)	5.0 NS	6.2*	8.0 NS
C (30–100 cases)	4.2 NS	6.2*	9.4*
D (> 100 cases)	7.6 ref	6.0 ref	4.7 ref
Length of stay	(Days)	(Days)	(Days)
A (1–9 cases)	1.7‡	2.0‡	2.1‡
B (10–29 cases)	1.6†	1.8‡	1.8*
C (30–100 cases)	1.5†	1.8‡	1.9‡
D (> 100 cases)	1.3 ref	1.5 ref	1.4 ref
Hospital charges	($)	($)	($)
A (1–9 cases)	3467 NS	4068 NS	4416‡
B (10–29 cases)	3499 NS	3696 NS	4046‡
C (30–100 cases)	3311†	3591†	3978 NS
D (> 100 cases)	3911 ref	4252 ref	3553 ref

Outcomes adjusted for surgeon volume and patient age, comorbidities, hospital volume and time period.
Key: NS = not significant; ref = reference category when dummy variables used to calculate p values in multivariate regression model.
* $p < 0.05$ compared to Group D (ref).
† $p < 0.01$ compared to Group D (ref).
‡ $p < 0.001$ compared to Group D (ref).
Ref = reference category when dummy variables used to calculate p values in multivariate regression model.

Table 6. ADJUSTED OUTCOMES OF SURGEON VOLUME GROUPS BY PROCEDURE

Outcome, by Surgeon Volume Group (Cases/yr)	Procedure			
	Lobectomy (n = 2705)	Other subtotal (n = 1766)	Substernal (n = 220)	Total (n = 1144)
Complications	(%)	(%)	(%)	(%)
A (1–9 cases)	7.7 NS	9.8*	18.8 NS	16.1‡
B (10–29 cases)	5.8 NS	5.9 NS	8.5 NS	11.7‡
C (30–100 cases)	5.6 NS	5.5 NS	16.6 NS	11.2‡
D (> 100 cases)	6.2 ref	6.6 ref	11.5 ref	4.3 ref
Length of stay	(Days)	(Days)	(Days)	(Days)
A (1–9 cases)	1.7‡	2.0‡	2.5*	2.4‡
B (10–29 cases)	1.6‡	1.7†	1.9 NS	2.0*
C (30–100 cases)	1.5†	1.8‡	2.1*	2.1†
D (> 100 cases)	1.3 ref	1.5 ref	1.8 ref	1.6 ref
Hospital charges	($)	($)	($)	($)
A (1–9 cases)	3652 NS	3808*	4676 NS	4866‡
B (10–29 cases)	3428‡	3549‡	3915*	4684*
C (30–100 cases)	3313‡	3393‡	4219 NS	4472 NS
D (> 100 cases)	3718 ref	4309 ref	4596 ref	4094 ref

Outcomes adjusted for surgeon volume and patient age, comorbidities, hospital volume, and time period.
Key: NS = not significant; ref = reference category when dummy variables used to calculate p values in multivariate regression model.
* $p < 0.05$ compared to Group D (ref).
† $p < 0.01$ compared to Group D (ref).
‡ $p < 0.001$ compared to Group D (ref).
Ref = reference category when dummy variables used to calculate p values in multivariate regression model.

hancing the quality, and reducing the cost, of surgical health care. If surgeons differ significantly in the outcomes of thyroidectomy as a function of their experience measured by caseload, directing more patients toward high-volume thyroid surgeons should result in fewer adverse surgical outcomes. The results of this study suggest that more than 20% of the complications and 1700 hospital days could have been saved in Maryland if all thyroidectomies were performed by high-volume surgeons.

There are several possible explanations for our finding of a relation between volume and outcomes. The "practice makes perfect" theory proposed by Luft et al.[26] argues that greater experience among higher-volume providers leads directly to better outcomes. The second explanation is that patients are more often referred to surgeons with better outcomes. In other words, referring physicians know which surgeons have better outcomes and selectively refer more patients to those surgeons.[18] It is impossible in this cross-sectional study to separate the influence of improved outcomes with practice. The third explanation is that lower-volume surgeons perform thyroidectomies on sicker patients, with resulting worse outcomes. This is a less likely explanation for the results of this study, because we used a measure of comorbidity for surgical risk adjustment.

One can only speculate as to why surgeon volume, rather than hospital volume, was the more important measure for thyroid procedures in this study. It is possible that an individual surgeon's experience is more critical for complicated, risky procedures such as thyroidectomy. At the same time, most patients with thyroid disease are relatively young and otherwise healthy. As a result, a superior patient outcome generally does not require a large perioperative team of surgeons, intensivists, and consultants, or complex hospital equipment and monitoring.

Kocher was the first high-volume endocrine surgeon, and his career provided early evidence supporting a relation between surgeon experience and clinical outcomes in thyroid surgery. During the first decade of his tenure in Berne, Kocher excised 101 goiters with a mortality rate approaching 13%; by 1889, he had performed 202 thyroidectomies with a mortality rate of 2.4%; by 1917, his personal series had swelled to more than 5000 thyroid operations with a mortality rate of 0.5%.[2]

There has been little objective evidence outside of institutional and personal series to show that surgeon volume is associated with thyroidectomy outcomes. An association between hospital volume and length of stay has been shown for parathyroidectomy, and a survey of endocrine surgeons showed that reported rates of complications and death after parathyroidectomy were associated with annual surgeon caseload.[27,28] Other studies have provided contradictory evidence assessing related factors, such as surgical training and specialty. Although there is evidence to suggest that otolaryngology residents can perform thyroidectomies safely when supervised by experienced faculty, there is also evidence to suggest that otolaryngologists have higher rates

Table 7. APPENDIX A. DETAIL OF COMPLICATIONS ICD-9-CM DIAGNOSIS AND PROCEDURE (P) CODES

Allergy/Drug Reaction	
99.52 (P)	Influenza Vaccination
E878.8	Abn React-Surg Proc NEC
E879.8	Abn React-Procedure NEC
Cardiovascular	
427.5	Cardiac Arrest
997.1	Surg Compl-Heart
997.91	Surg Comp-Hypertension
Endocrine	
252.1	Hypoparathyroidism
Gastrointestinal	
997.4	Surg Comp-Digestv System
Hematologic/Vascular	
39.98 (P)	Hemorrhage Control NOS
99.04 (P)	Packed Cell Transfusion
996.74	Comp-Oth Vasc Dev/Graft
998.1	Hemorr Complic Procedure
998.2	Accidental Op Laceration
E870.0	Acc Cut/Hem in Surgery
Neurologic	
478.3	Paral Vocal Cord/Larynx
478.30	Vocal Cord Paralysis NOS
478.31	Vocal Paral Unilat Part
478.32	Vocal Paral Unilat Total
478.33	Vocal Paral Bilat Part
478.34	Vocal Paral Bilat Total
997.0	Surg Complication-CNS
997.00	Nervous Syst Complc NOS
997.09	Surg Comp Nerv Systm NEC
Urologic	
599.0	Urin Tract Infection NOS
996.31	Malfunc Urethral Cath
996.76	Comp-Genitourin Dev/Grft
997.5	Surg Compl-Urinary Tract
Respiratory	
462	Acute Pharyngitis
518.4	Acute Lung Edema NOS
518.5	Post Traum Pulm Insuffic
519.0	Tracheostomy Complic
96.72 (P)	Cont Mech Vent 96+ Hrs
997.3	Surg Complic-Respir Syst
Wound	
06.01 (P)	Thyroid Field Aspiration
06.02 (P)	Reopen Thyroid Field Wnd
682.1	Cellulitis of Neck
86.04 (P)	Other Skin & Subq I & D
86.22 (P)	Exc Wound Debridement
998.3	Postop Wound Disruption
998.5	Postoperative Infection
Other (Abnormal Reactions)	
996.59	Malfunc Oth Device/Graft
997.9	Surg Compl-Body Syst NEC
998.8	Surgical Complicat NEC
998.89	Oth Spcf Cmplc Procd NEC
998.9	Surgical Complicat NOS
999.9	Complic Med Care NEC/NOS
E878	Abn Reaction to Surgery
E878.4	Abn React-Plast Surg NEC
E878.6	Abn Reac-Organ Rem NEC

NEC = not elsewhere classified; NOS = not otherwise specified.

of permanent postsurgical hypoparathyroidism than their general surgery colleagues.[20,29]

Variation in practice patterns in the surgical management of thyroid disease makes measuring outcomes from thyroidectomy especially complex. The optimal extent of thyroid gland removal has been widely debated, with opponents of extensive resection arguing that these procedures carry potentially increased complications. Advocates of total thyroidectomy for cancer maintain that the procedure decreases recurrence, facilitates ^{131}I treatment, improves survival, and is associated with only a slight increase in risk. As a result of conflicting evidence, considerable variation exists even among surgeons with expertise in endocrine surgery about the optimal surgical management of thyroid disease.[30] We adjusted separately in our models for thyroid diagnosis and procedure, and this allowed us to account for variation in surgeons' choice of procedure within diagnosis categories.

This study has several possible limitations. There may be some coding errors of diagnoses or procedures in the HSCRC database. In addition, without longitudinal data, it is difficult to distinguish between complications and comorbidities.[31] We were able to check the accuracy of reporting for complications against a medical chart review for 8% of the study population, all of whom underwent thyroidectomy at the high-volume hospital. There is no reason to suspect that the accuracy of reporting of our diagnoses of interest would be different across hospitals. Overall, complication rates reported in this study are similar to those in the published literature.[32] It is important to emphasize that our database captured all perioperative complications, including those that were not directly related to the procedure. As a result, they appear to be higher than in studies that report only rates of recurrent laryngeal nerve injury, hypoparathyroidism, and wound complications. However, they are probably a conservative estimate of adverse patient events overall because they exclude some complications detected after hospital discharge, such as hypoparathyroidism and vocal cord paralysis.

This study provides compelling evidence for a significant association between increased surgeon volume and improved patient outcomes after surgical procedures for both benign and malignant thyroid disease. This evidence supports the referral of patients for thyroidectomy to high-volume surgeons. The relation between surgeon experience and complication rates, length of stay, and hospital charges is strongest for more complex thyroid diagnoses and procedures. Therefore, the case for referral to high-volume surgeons is especially compelling for patients with known or suspected thyroid cancer or patients for whom a near-total or total thyroidectomy is the procedure of choice. These findings have implications for patients, referring physicians, medical educators, and third-party payers responsible for setting medical coverage and reimbursement policies.

References

1. Tunbridge WMG, Evered DC, Hall R, et al. The spectrum of thyroid disease in a community: the Whickham survey. Clin Endocrinol 1977; 7:481–493.
2. Becker WF. Pioneers in thyroid surgery. Ann Surg 1977; 185:493–504.
3. Kocher T. Zur pathologie und therapie des kropfes. Dtsch Z Chirug 1874; 4:417.
4. Halsted WS. The operative story of goiter. The author's operation. Johns Hopkins Rep 1920; 19:71.
5. Lahey FH. Routine dissection and demonstration of the recurrent laryngeal nerve in subtotal thyroidectomy. Surg Gynecol Obstet 1938; 66:775.
6. Crile GW. The prevention of abductor paralysis in thyroidectomy. Surg Gynecol Obstet 1929; 49:538.
7. Lore JM Jr. Surgery of the thyroid gland. Otolaryngol Clin North Am 1980; 13(1):69–83.
8. Edis AJ. Prevention and management of complications associated with thyroid and parathyroid surgery. Surg Clin North Am 1979; 59(1):83–92.
9. Farrar WB. Complications of thyroidectomy. Surg Clin North Am 1983; 63(6):1353–1361.
10. Harness JK, Fung L, Thompson NW, et al. Total thyroidectomy: complications and technique. World J Surg 1986; 10:781–786.
11. Shaha A, Jaffe BM. Complications of thyroid surgery performed by residents. Surgery 1988; 104:1109–1114.
12. Gordon TA, Burleyson GP, Tielsch JM, Cameron JL. The effects of regionalization on cost and outcomes for one general high-risk surgical procedure. Ann Surg 1995; 221(1):43–49.
13. Lieberman MD, Kilburn MA, Lindsey M, Brennan MF. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995; 222(5):638–645.
14. Hannan EL, O'Donnell JF, Kilburn H, et al. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. JAMA 1989; 262(4):503–510.
15. Adams DF, Fraser DB, Abrams HL. The complications of coronary arteriography. Circulation 1973; 48:609.
16. Jollis JG, Peterson ED, DeLong ER, et al. The relation between the volume of coronary angioplasty procedures at hospitals treating Medicare beneficiaries and short-term mortality. N Engl J Med 1994; 331(24):1625–1629.
17. Lavernia CJ, Guzmann JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty 1995; 10(2):133–140.
18. Harris SC. Thyroid and parathyroid surgical complications. Am J Surg 1992; 163:476–478.
19. Martin L, Delbridge L, Martin J, et al. Trainee surgery in teaching hospitals: is there a cost? Aust NZ J Surg 1989; 59:257–260.
20. Shindo ML, Sinha UK, Rice DH. Safety of thyroidectomy in residency: a review of 186 consecutive cases. Laryngoscope 1995; 105:1173–1175.
21. Reeve TS, Curtin A, Fingleton L, et al. Can total thyroidectomy be performed as safely by general surgeons in provincial centers as by surgeons in specialized endocrine surgical units? Arch Surg 1994; 129:834–836.
22. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993; 46(10):1075–1079.
23. Charlson ME, Pompei P, Alex KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987; 40(5):373–383.
24. Health Care Financing Administration, OACT, Office of National Health Statistics. DRI/McGraw-Hill HCC, 1995.
25. Health Services Cost Review Commission annual cost filings and approved rate orders, FY1990–FY1995.
26. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 1979; 301:1364–1369.
27. Chen H, Zeiger MA, Gordon TA, Udelsman R. Parathyroidectomy in Maryland: effects of an endocrine center. Surgery 1996; 120(6):948–953.
28. Sosa JA, Powe NR, Levine MA, et al. Thresholds for surgery and surgical outcomes for patients with primary hyperparathyroidism: a national survey of endocrine surgeons. J Clin Endocrinol Metab (1998, in press).
29. Burge MR, Zeise T-M, Johnsen MW, et al. Risks of complication following thyroidectomy. J Gen Intern Med 1998; 13:24–31.
30. Harrison BJ, Wheeler MH. Review of current practice in surgical management of benign thyroid disease. Br J Surg 1993; 80:1209.
31. Roos LL, Stranc L, James RC, Li J. Complications, comorbidities, and mortality: improving classification and prediction. Health Serv Res 1997; 32(2):229–238.
32. Van Heerden JA, Groh MA, Grant CS. Early postoperative morbidity after surgical treatment of thyroid carcinoma. Surgery 1987; 101(2): 224–227.

Discussion

DR. SAMUEL A. WELLS, JR. (St. Louis, Missouri): I have two concerns about their study: When endocrinologists and endocrine surgeons think of complications associated with thyroidectomy they primarily refer to nerve injury (either the recurrent laryngeal nerve or the external branch of the superior laryngeal nerve) and hypoparathyroidism. In the present study all operative complications, including drug reactions, wound infections, and inflammation of the pulmonary tree and the urinary tract, were grouped with nerve injury and hypoparathyroidism, and compared among the four groups of surgeons. The overall incidence of complications was higher in the more experienced surgeons compared to the least experienced and surgeons who performed the most thyroidectomies had fewer nerve injuries than surgeons who performed the least. However, the incidence of hypoparathyroidism in patients was the same among the four surgical groups and there was no relationship to operative experience. Do the authors have an explanation for this?

The recommendation that the difficult thyroid cases be centralized in the high volume centers seems impractical since the diagnosis of thyroid cancer, or the decision to perform a total thyroidectomy, may only become apparent intraoperatively. Furthermore, although not mentioned in the text, it is likely that an informal referral system was already in play with patients being referred to experienced surgeons both from their surgical colleagues and from endocrinologists.

Also, the 78% of surgeons who performed the fewest number of thyroidectomies probably performed other operations infrequently, and possibly had more complications than their surgical specialist colleagues in large referral centers who perform a large number of the given operative procedures. The surgical generalist while not having an inordinately large experience with a single operative procedure may be a superior surgeon overall than the surgical specialist and he or she provides an invaluable service to his or her medical community. The best surgeons, regardless of their practice setting, know their limitations and refer complicated or challenging cases which require the care of a physician or surgeon who is highly experienced and knowledgeable in the area of interest.