PAPER

Are Young Surgeons Competent to Perform Alimentary Tract Surgery?

Jay B. Prystowsky, MD, MHPE

Background: Assessment of competency during residency training has received increased attention recently. There has been less attention given to the competency of residents after training.

Hypothesis: Patient outcomes for alimentary tract surgery (ATS) should be similar for surgeons who recently completed their residency training compared with more experienced surgeons, indicating that the younger surgeons had achieved clinical competency on completion of their residency training.

Design: Retrospective analysis of Illinois inpatient discharge data (January 1, 1996–December 31, 1999).

Setting: All 205 nonfederal acute care hospitals in Illinois.

Patients: The patients were 120 160 adult Illinois residents who underwent ATS in Illinois.

Main Outcome Measures: Mortality rate, morbidity rate, and hospital length of stay.

Results: Regression analyses demonstrated that surgeon experience was a significant determinant of mortality and morbidity rates, with worse outcomes observed for patients of young surgeons undergoing high-complexity ATS (ie, procedures other than appendectomy and cholecystectomy).

Conclusions: For high-complexity ATS, there was a significant disparity in outcomes between young and more experienced surgeons, whereas for low-complexity ATS, there was no disparity. Attention to competency during residency training is warranted, especially as it relates to high-complexity ATS. Furthermore, patient outcomes provide an opportunity to assess competency after training that can complement assessments during training and together identify educational strengths and weaknesses of residency training.

Arch Surg. 2005;140:495-502

Author Affiliation: Department of Surgery, Northwestern University Feinberg School of Medicine, Chicago, Ill.

"The ultimate goal of every surgical training program is to produce competent professionals capable of meeting the healthcare needs of society."[1(p6)]

THE PREVIOUSLY MENTIONED statement succinctly captures the essence of surgical residency training. In 1999, the Accreditation Council for Graduate Medical Education in the United States defined the attributes that competent physicians must possess to discharge their duties as physicians.[2] The Accreditation Council for Graduate Medical Education identified 6 general competencies that represent the skill sets that all physicians should possess: patient care, medical knowledge, practice-based learning, interpersonal and communication skills, professionalism, and systems-based practice. Recently, the American Board of Surgery (ABS) embraced these competencies.

Residency programs are responsible for implementing a curriculum that addresses these competencies and a plan for their evaluation. Training programs must identify educational outcomes and objectives, and determine how well their residents are attaining these objectives. There are numerous benefits to evaluating educational outcomes. For individual residents, systematic feedback of strengths and weaknesses is of paramount importance in their training process. In addition, the residency program may benefit as a whole in that performance of its residents may help to identify its own strengths and weaknesses in curriculum, instruction, and evaluation.

The assessment tools to evaluate competency are diverse.[1] Ideally, assessment should be reliable, valid, and practical. Within a training program, there are many methods for residency assessment, and some examples include oral and written examinations, 360° reviews, and performance-based assessments. To some extent, all of these methods have

Table 1. *ICD-9-CM* **Codes for 120 160 Alimentary Tract Operations**

Organ System	*ICD-9-CM* Code	No. (%) of Operations Performed
Esophagus	>423.9 and <430.0	577 (0.5)
Stomach	>429.9 and <450.0	9114 (7.6)
Small and/or large intestine	>449.9 and <470.0	39 432 (32.8)
Appendix	>469.9 and <480.0	19 113 (15.9)
Rectum	>479.9 and <490.0	4189 (3.5)
Liver	>499.9 and <510.0	769 (0.6)
Bile duct	>512.9 and <520.0	1309 (1.1)
Gallbladder	>509.9 and <512.9	39 945 (33.2)
Pancreas	>519.9 and <530.0	1376 (1.1)
Laparotomy	>540.0 and <550.0	4336 (3.6)

Abbreviation: *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification.*

limitations in terms of being an ideal assessment tool, but nevertheless, they are generally useful.

As previously stated, the critical goal of a training program is to produce competent physicians who are able to provide quality care to patients. The effectiveness of that process is generally measured by assessment of trainee performance while the trainee is matriculating in the educational program. Assessments in residency may be useful, but at best, they are surrogates for educators who are trying to determine if their trainees will work safely and competently in their practice after residency. There is little knowledge of how residents perform as physicians in the community.[3] Yet, that is exactly the role for which training programs prepare them. Measurement of physician performance is challenging, but the growth and development of health care outcomes research during the past 2 decades provides helpful insights and methods for the assessment of physician performance.[4]

The purpose of this investigation was to determine if patient outcomes for alimentary tract surgery (ATS) varied according to surgeon experience and, specifically, to determine if patient outcomes were worse for surgeons who recently completed their residency training. In this manner, the strengths and weaknesses of training programs might be identified in terms of their ability to produce competent physicians who are able to meet prevailing community standards of quality health care as it relates to ATS.

METHODS

Institutional review board approval for conduct of this study was obtained from Northwestern University Feinberg School of Medicine.

DATA

Administrative data on Illinois residents discharged from nonfederal Illinois hospitals from January 1, 1996, through December 31, 1999, were obtained from the Illinois Hospitals and Health Systems Association Compdata files. These files contain publicly mandated discharge information for Illinois residents from Illinois hospitals, and include age, sex, hospital, type of admission, vital status at discharge, principal procedure, primary diagnosis, and up to 8 secondary diagnoses for each admission (diagnostic codes are derived from the *International Classification of Diseases, Ninth Revision, Clinical Modification* manual for preexisting diseases and inpatient complications).[5] Admissions of adult patients (≥18 years old) for whom an operation of the alimentary tract was the primary procedure were selected for analysis (**Table 1**). Alimentary tract procedures included operations on the esophagus, stomach, small and large intestine, appendix, rectum, liver, pancreas, gallbladder, and bile ducts. Primary procedures that were listed as laparotomies were also included. Patients whose primary procedure was upper or lower endoscopy or an anal procedure were not included. Patient variables were age, type of admission, and comorbid illnesses. Categorization of age (18-40, 41-60, 61-80, and ≥81 years) was performed for ease of interpretation and was based on clinical norms.[6] Type of admission was modeled as a dichotomous variable, namely, emergency or nonemergency admission. Emergency admission was defined purely based on the coding in the discharge information submitted to the state.

PATIENT RISK STRATIFICATION

Twenty-seven comorbid illnesses were identified from several previous analyses of comorbid conditions derived from administrative databases or surgical outcomes' studies (**Table 2**).[7-13] These analyses were performed to evaluate severity of illness in populations of surgical and medical patients. Each illness was modeled as a dichotomous variable, and χ^2 analysis with mortality was performed. Logistic regression analysis of all illnesses for mortality was performed to determine the relative contribution of comorbid illnesses to mortality. Odds ratios for mortality of greater than 1.5 were used to select those comorbid illnesses that most adversely affected mortality, and the coefficients for each illness were rounded to the nearest integer and used to develop a weighted index of severity of illness for each patient. The severity of illness index permitted grouping of patients into low, medium, and high severity of illness based on the type and number of their comorbid illnesses. In this manner, severity of illness was converted into a categorical variable. In addition, a C statistic (area under a receiver operating characteristic curve) for mortality was calculated to determine how well this formula discriminated between survivors and nonsurvivors.[14] A C statistic of 0.5 suggests no predictive value, while 1.0 suggests perfect predictive value. A C statistic of approximately 0.8 or greater suggests a model of excellent predictive value.

SURGEON CLASSIFICATION

The Compdata files include information on attending physicians or surgeons. The American Board of Medical Specialties' directories from 1997, 1998, and 2000, of certified physicians, were used to confirm the status of individual physicians certified by the ABS.[15-17] Surgeon experience was classified categorically according to number of years since ABS certification. Four groups were defined according to the year in which ABS certification was obtained: 1994 to 2001, 1979 to 1993, 1969 to 1978, and before 1969. Thus, no surgeon from the first group would have more than 5 years' experience during the study period of 1996 to 1999. The maximum years of experience in the next 2 groups would, therefore, be 20 and 30 years, respectively, and surgeons in the last group had at least 31 years' experience, as defined by years since ABS certification. Patients with no identifiable surgeon or whose surgeon was not board eligible or board certified according to discharge information were excluded from analysis.

Table 2. *ICD-9-CM* Codes for Comorbid Illnesses

Comorbid Illness	*ICD-9-CM* Code	Data for Mortality	
		P Value*	OR (95% CI)
AIDS	>419.0 and <450.0	<.05	3.7 (2.9-4.6)
Alcohol and/or other drug dependence	>302.9 and <305.0	<.05	1.3 (1.0-1.6)
Anemia	>280.0 and <285.1 and >285.1 and <286.0	NS	0.7 (0.6-0.8)
Central nervous system disease	>319.9 and <342.0 and >344.9 and <346.0	<.05	1.7 (1.4-2.1)
Cerebrovascular disease	>429.9 and <439.0	<.05	2.5 (2.1-2.9)
Chronic obstructive pulmonary diseases	>489.9 and <500.0	<.05	1.3 (1.1-1.4)
Chronic liver disease	571.0, >571.1 and <572.0, and >572.1 and <572.9	<.05	2.4 (2.1-2.8)
Coagulopathy	>285.9 and <287.0, 287.1, and >287.2 and <288.0	<.05	7.2 (6.5-7.9)
Congestive heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, >415.9 and <417.0, and >427.9 and <429.0	<.05	4.8 (4.5-5.2)
Diabetes mellitus			
Uncomplicated	>246.9 and <250.1	<.05	0.8 (0.7-0.9)
Complicated	>250.0 and <251.0	<.05	1.4 (1.2-1.6)
Hypertension			
Uncomplicated	>401.0 and <402.0	NS	0.6 (0.5-0.6)
Complicated	>401.9 and <410.0	<.05	1.0 (0.9-1.3)
Leukemia	>203.8 and <210.0	<.05	3.8 (2.8-5.2)
Lymph node metastasis	>195.9 and <197.0	<.05	0.9 (0.8-1.0)
Lymphoma	>199.1 and <202.1	<.05	1.6 (1.0-2.6)
GI tract malignancy	>149.9 and <160.0	<.05	1.1 (1.0-1.2)
Malnutrition	>259.0 and <264.0	<.05	2.6 (2.3-2.9)
Metastatic disease	>196.9 and <200.0	<.05	2.9 (2.6-3.1)
Non–GI tract malignancy (excluding skin)	>139.9 and <150.0, >159.9 and <173.0, and >173.9 and <196.0	<.05	1.9 (1.7-2.3)
Obesity	>277.9 and <278.1	NS	0.4 (0.3-0.5)
Paraplegia or quadriplegia	>341.9 and <345.0	<.05	0.9 (0.7-1.3)
Peripheral vascular disease	>440.1 and <441.0 and >442.9 and <444.0	<.05	1.3 (1.0-1.6)
Psychoses	>289.9 and <300.0	<.05	1.0 (0.9-1.2)
Renal failure	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 250.4, and >584.9 and <587.0	<.05	3.5 (2.8-4.5)
Rheumatic and connective tissue disorders	>709.9 and <711.0 and >713.9 and <715.0	<.05	1.4 (1.1-1.8)
Valvular heart disease	>093.19 and <093.25, >393.9 and <397.2, and >746.2 and <746.7	<.05	1.1 (0.9-1.5)

Abbreviations: CI, confidence interval; GI, gastrointestinal; *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification*; NS, not signigicant; OR, odds ratio.

*Data obtained using the χ^2 test.

Surgeon volume was classified as a continuous variable. However, patients were included only if their surgeon had performed more than 10 operations in the categories of either low-complexity operations (appendectomy and cholecystectomy) or high-complexity operations (all other procedures) during the study period. In this manner, patient outcomes of surgeons who perform operations rarely in either category would be excluded from analysis.

MAIN OUTCOME MEASURES

Main outcome measures were mortality, morbidity, and hospital length of stay (LOS). Inpatient postoperative complications were identified in the following categories: cardiac, digestive, iatrogenic, respiratory, thrombosis, urinary tract, and wound infection (**Table 3**). Patients were assigned to the complication group if they had at least 1 inpatient complication. Postoperative morbidity and mortality were treated as dichotomous variables. The LOS (in days) was treated as a continuous variable.

STATISTICAL ANALYSIS

Descriptive statistics were used to describe characteristics of the patient population and surgeons. Preliminary analysis of the main outcomes according to type of operation demonstrated that outcomes were better for low-complexity procedures (appendectomy and cholecystectomy) compared with all other procedures, using χ^2 analysis for mortality and morbidity and an independent-sample *t* test for LOS. Subsequent analyses were performed separately for low- and high-complexity operations.

χ^2 Analysis was used to assess the relationship of surgeon experience with mortality and morbidity. An independent-sample *t* test was used to assess the relationship of surgeon experience with LOS. Logistic regression was used to assess the significance of surgeon experience on mortality and morbidity, and linear regression was used to assess the significance of surgeon experience on LOS after controlling for patient age, admission type, severity of illness, and surgeon volume.

RESULTS

From 1996 to 1999, 170 769 Illinois residents (aged ≥18 years) underwent ATS at 205 nonfederal acute care Illinois hospitals. Among this group of patients, there were 120 160 admissions (70.4%) for which the surgeon could be clearly identified. Subsequent analyses and discussion refer to this cohort of admissions.

The mean ± SD patient age was 56 ± 20 years (range, 18-106 years). The female-male ratio was 57:43. Emer-

Table 3. *ICD-9-CM* Codes for Postoperative Complications

Category	Diagnosis	*ICD-9-CM* Code
Cardiac	Acute myocardial infarction	>409.9 and <411.0
	Cor pulmonale	415.0
	Acute left ventricular failure	428.1
Digestive	Complications of intestinal anastomosis or bypass and intestinal obstruction due to a procedure	997.4
Iatrogenic	Postoperative hemorrhage or hematoma, unintentional laceration or puncture, wound dehiscence, and foreign body left in wound unintentionally	>998.0 and <998.5
	Iatrogenic pneumothorax	512.1
Respiratory	Respiratory complication	997.3
	Bacterial pneumonia	>480.9 and <483.0
	Aspiration pneumonitis	507.0
	Pulmonary edema (noncardiogenic)	514.0
	Respiratory failure or adult respiratory distress syndrome	>518.3 and <519.0
Thrombosis	Venous thrombosis (superficial or deep, at any location)	>450.0 and <454.0
	Pulmonary embolism	415.1
Urinary tract	Urinary tract infection	599.0
	Urinary tract obstruction	599.6
	Urinary tract complication	>997.4 and <997.6
	Acute renal failure	>583.9 and <585.0
Wound	Postoperative wound infection	>998.4 and <999.8

Abbreviation: *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification*.

Table 4. Clinical Outcomes According to Surgeon Experience

Time Since ABS Certification, y	Mortality Rate, %	Morbidity Rate, %	LOS, d*
Low-Complexity Operations			
0-5	0.5	12.8	4.1 ± 4.3
6-20	0.6	13.4	4.2 ± 4.3
21-30	0.6	12.6	4.4 ± 4.9†
>30	0.6	13.6	4.6 ± 4.3†
High-Complexity Operations			
0-5	7.0	34.0	10.4 ± 9.2
6-20	5.2†	30.4†	10.3 ± 8.7
21-30	5.1†	30.0†	10.6 ± 8.4
>30	5.8†	30.4†	11.0 ± 9.0†

Abbreviations: ABS, American Board of Surgery; LOS, length of stay.
*Data are given as mean ± SD.
†$P<.05$ vs the 0 to 5 years group, using regression analysis.

gency admissions occurred in 39.8% of cases. The overall mortality was 3.0%. The overall morbidity was 21.9%. The incidence of complications by category was as follows: cardiac, 2.7%; digestive, 6.4%; iatrogenic, 3.6%; respiratory, 6.5%; thrombosis, 0.8%; urinary tract, 6.1%; and wound, 2.4%. The mean±SD hospital LOS was 7.4±7.5 days (range, 1-327 days).

There were 24 illnesses that were significantly associated with mortality by χ^2 analysis (Table 2), of which 12 were associated with an odds ratio for mortality of greater than 1.5. Of these illnesses, 7 were associated with a 2- to 3-fold increase in the risk of death (type 1 illness: central nervous system disease, cerebrovascular disease, cirrhosis, lymphoma, malnutrition, metastatic disease, and non–gastrointestinal tract malignancy) and 5 were associated with at least a 4-fold increase in the risk of death (type 2 illness: AIDS, coagulopathy, congestive heart failure, leukemia, and renal failure). These 12 illnesses were subsequently used to classify patients according to severity of illness: low severity indicates no illness; medium severity, 1 type 1 illness; and high severity, at least 2 type 1 illnesses or 1 type 2 illness. The percentages of all patients who were assigned to each category were 80.8%, 9.6%, and 9.6%, respectively. Overall, the C statistic for survival was 0.80 using this formula for severity of illness, indicating that it is a satisfactory method for risk stratification based on comorbid illnesses.

The percentages of operations performed by each group of surgeons according to experience were as follows: group 1 (0-5 years), 8.8%; group 2 (6-20 years), 59.4%; group 3 (21-30 years), 26.2%; and group 4 (≥31 years), 5.6%. The percentage of procedures in each category of operations is given in Table 1. There was a significant difference in clinical outcomes when comparing the results of low-complexity operations (n=59 058 [49.1%]; appendix and gallbladder) with those of high-complexity operations (n=61 102 [50.9%]; all other procedures): 0.6% vs 5.4% for mortality, 13.3% vs 31.2% for morbidity, and a mean±SD of 4.3±4.6 vs 10.4±8.7 days for LOS ($P<.05$ for all).

Outcomes according to experience are listed in **Table 4**. χ^2 Analysis demonstrated a significant relationship of less experienced surgeons to increased mortality and morbidity ($P<.05$), but there was no significant relationship to LOS. Regression analyses demonstrated increasing age, emergency admission, and increasing severity of illness to have a negative influence on outcomes for low- and high-complexity operations. There was no significant influence of surgeon experience for outcomes of patients undergoing low-complexity operations, but there was a significant positive influence of greater surgeon experience on mortality and morbidity for patients undergoing high-complexity operations (**Table 5**). Therefore, even after controlling for patient age, emergency admission, severity of illness, and surgeon volume, patient outcomes for high-complexity ATS were worse for surgeons with 5 years' experience or less since ABS certification.

COMMENT

The focus of this investigation was patient outcomes of surgeons who had achieved ABS certification within 5 years of the study period. The results of this study suggest that this group of surgeons can safely perform appendectomies and cholecystectomies, as evidenced by similar patient mortality, morbidity, and hospital LOS when compared with the results of the more experienced surgeons. This finding implies that training programs are adequately preparing their trainees to perform these operations safely in the community.

However, there was a disparity in patient outcomes for high-complexity procedures. Given that high-complexity procedures occur less frequently than low-complexity procedures, it is not a complete surprise that there is a disparity in patient outcomes between less and more experienced surgeons in that the former group may simply not have performed sufficient numbers of these procedures to make them as safe as those who have added clinical experience once in practice.

By definition, the nature of high-complexity procedures makes them more challenging, and the learning curve to achieve proficiency may be lengthier. To enlighten this discussion, it is instructive to review several literature reports regarding learning curves for various operations.

Schauer et al[18] studied the learning curve for laparoscopic Roux-en-Y gastric bypass in their first 150 consecutive patients. They found a 50% reduction in technical complications after their first 100 patients in addition to a substantial decrease in operative time. Kligman et al[19] studied their experience with laparoscopic Roux-en-Y gastric bypass in 160 patients. They also observed a significant decrease in operative time after 120 patients, but did not observe a difference in complication rate between their early and late patients. Flum and Dellinger[20] performed a population-based analysis on the impact of gastric bypass on survival. They studied early and late mortality in 3328 patients. They observed a 30-day mortality of 1.9%. Within the surgeon's first 19 procedures, the odds ratio for 30-day mortality was 4.7 times higher (95% confidence interval, 1.2-18.2) than at later points in a surgeon's case order. McMasters et al[21] studied sentinel lymph node biopsy in 2148 patients operated on by 226 surgeons. They found that improvement in sentinel lymph node identification and reduction in false-negative rates occurred after the surgeon's first 20 cases. Bennett et al[22] evaluated the learning curve for laparoscopic-assisted colectomy. They studied outcomes of 1194 patients who were operated on by 114 surgeons. They found no difference in LOS and the conversion rate, but using regression analysis, they found significantly reduced intraoperative and postoperative complication rates for surgeons who had performed 40 or more procedures. They concluded that there is a learning curve for this procedure that manifests itself in patient complication rates. Watson et al[23] studied the learning curve for laparoscopic Nissen fundoplication. They evaluated outcomes in 280 patients operated on by 11 surgeons. They observed higher complication, reoperation, and conversion rates in each individual's first 20 procedures, with the highest complication rates in each individual's first 5 procedures. Flum et al[24] performed a nationwide population-based study of adverse outcomes for antireflux surgery. They studied mortality, splenectomy, and esophageal injury in 86 411 patients. The odds ratios of these adverse events were 5.6, 2.7, and 2.3 times greater, respectively, if the surgeon was performing one of his or her first 15 procedures compared with a later procedure. Moffat et al[25] investigated the learning curve for the removal of vestibular schwannomas in 300 patients, and observed satisfactory outcomes in 52% of the first 50 procedures, which improved significantly to 78% in the next 50 procedures, and was up to 92% in the last 50 procedures. Cochrane and Kestle[26] studied patient outcomes of ventriculoperitoneal shunt insertions in 3794 patients, performed by 254 surgeons during a 12-year period. Surgical experience was represented by the number of shunt insertions performed by each surgeon before the date of the operation. They observed significant decreases in the infection rate and the shunt failure rate with increasing experience.

Table 5. Regression Analyses of Clinical Outcomes for 60 591 High-Complexity Operations

Variable*	Mortality†	Morbidity†	B (SE) for Length of Stay‡
Age, y			
41-60	1.4 (1.1-1.6)§	1.4 (1.3-1.5)§	1.3 (0.1)§
61-80	2.5 (2.1-3.1)§	1.8 (1.7-1.9)§	2.3 (0.1)§
>80	4.0 (3.3-4.8)§	2.0 (1.9-2.2)§	2.2 (0.1)§
Emergency admission	2.5 (2.4-2.7)§	1.6 (1.5-1.6)§	3.5 (0.1)§
Risk severity of illness			
Medium	1.9 (1.7-2.2)§	1.4 (1.4-1.5)§	2.3 (0.1)§
High	6.7 (6.0-7.5)§	2.4 (2.3-2.5)§	5.8 (0.1)§
Surgeon experience, y			
6-20	0.8 (0.7-0.9)§	0.9 (0.8-1.0)§	0.1 (0.1)
21-30	0.7 (0.6-0.8)§	0.8 (0.8-0.9)§	0.2 (0.1)
>30	0.8 (0.6-0.9)§	0.8 (0.8-0.9)§	0.6 (0.2)§

*The reference variables were as follows: age of 18 to 40 years, nonemergency admission, low-risk severity of illness, and surgeon experience of 0 to 5 years.
†Data are given as odds ratio (95% confidence interval).
‡The constant for B (the regression coefficient for linear regression analysis) was 5.9 (SE, 0.2) ($P<.05$).
§$P<.05$.

The predominant message of these studies is that surgical learning curves exist, they are procedure specific, and they can be measured. In general, patient complications will result early in the learning curve without adequate supervision. In addition, for these high-complexity procedures, the learning curve typically ranged from 15 to 50 procedures.

The adage that "there is no substitute for experience" is true. However, it is not clear that training programs are providing sufficient experience in certain categories of operations to permit trainees to successfully negotiate the learning curves for those procedures. We take it for granted that ABS certification ensures a satisfactory level of surgical proficiency, and ATS falls squarely within the purview of ABS certification. However, the findings of this study suggest that a trainee may be ABS certified, but not proficient in certain alimentary tract procedures. In short, ABS certification does not ensure competence across the spectrum of the alimentary tract. Hospital credentialing typically takes ABS certification at face value and permits credentialed surgeons to operate throughout the alimentary tract. However, the operative experience of a resident may not necessarily conform to the learning curve for a given procedure. Furthermore, it is not clear that the minimum number of cases suggested by the ABS is based on evidence related to the learning curve for particular procedures. Therefore, the public is left to the self-judgment of surgeons as to whether they are proficient at a procedure and, if not, whether they have sought adequate supervision subsequent to residency training. The results of this study suggest that many less experienced surgeons are not prepared to perform high-complexity ATS at acceptable complication rates. Accordingly, either additional experience during residency or additional supervision after residency is warranted.

The health care outcomes' initiatives of the past 2 decades provide a framework for measuring outcomes and will further facilitate the measurement of physician performance and establishment of best practices in the community. Surgical investigators have studied the learning curve for various procedures, and their results suggest that evidence-based guidelines can be developed for achieving technical proficiency for specific operations. Naturally, becoming a competent surgeon is more than performing a certain minimum number of select procedures. However, there is growing evidence that suggests a specified number of procedures to attain technical proficiency for an operation. The corollary is that residency programs may be well served to use evidence related to learning curves for procedures, to guide their decision making in establishing the operative experience of their trainees. In addition, graduating residents may want to use this information as a guide to seeking supervision or assistance as they establish their own surgical practice in the community.

As residency programs struggle with assessment of the 6 general competencies of their trainees, it is worth remembering that the ultimate goal of training is to prepare residents to provide quality health care in the community. It is important to assess competency during residency, but it is also worthwhile to assess competency after training, as challenging as that might be. As health care outcomes' research advances, the opportunity to study patient outcomes in the community will be facilitated and the information that results will be useful feedback to practicing physicians and the residency programs in which they trained.

Accepted for Publication: January 18, 2005.
Correspondence: Jay B. Prystowsky, MD, MHPE, Department of Surgery, Northwestern University Feinberg School of Medicine, 201 E Huron St, Galter 10-105, Chicago, IL 60611 (jprystowsky@nmff.org).
Previous Presentation: This paper was presented at the Annual Meeting of the Western Surgical Association; November 9, 2004; Las Vegas, Nev; and is published after peer review and revision. The discussions that follow this article are based on the originally submitted manuscript and not the revised manuscript.

REFERENCES

1. Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? *Surgery*. 2004;135:6-20.
2. Accreditation Council for Graduate Medical Education. *General Competencies: Version 1.3: September 28, 1999*. Chicago, Ill: Accreditation Council for Graduate Medical Education; 2000.
3. Prystowsky JB, Bordage G. An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ*. 2001;35:331-336.
4. Slater C. What is outcomes research and what can it tell us? *Eval Health Prof*. 1997;20:243-264.
5. *International Classification of Diseases, Ninth Revision, Clinical Modification*. Washington, DC: Public Health Service, US Dept of Health and Human Services; 1988.
6. Showstack J, Rosenfeld K, Garnick D, Luft H, Scharrarzick R, Fowles J. Association of volume with outcome of coronary artery bypass graft surgery. *JAMA*. 1987;257:785-789.
7. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method for classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373-383.
8. 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;46:1075-1079.
9. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with *ICD-9-CM* administrative databases. *J Clin Epidemiol*. 1992;45:613-619.
10. Roos LL, Fisher ES, Sharp SM, et al. Postsurgical mortality in Manitoba and New England. *JAMA*. 1990;263:2453-2458.
11. Roos LL, Stranc L, James RC, Li J. Complications, comorbidities, and mortality: improving classification and prediction. *Health Serv Res*. 1997;32:229-242.
12. Khuri SF, Daley J, Henderson W, et al. Relation of surgical volume to outcome in eight common operations. *Ann Surg*. 1999;230:414-432.
13. Pronovost P, Dorman T, Sadovnikoff N, et al. The association between preoperative patient characteristics and both clinical and economic outcomes after abdominal aortic surgery. *J Cardiothorac Vasc Anesth*. 1999;13:549-554.
14. Royston P. An extension of Shapiro and Wilk's W test for normality to large samples. *Appl Stat*. 1982;31:115-124.
15. American Board of Medical Specialties. *The Official ABMS Directory of Board Certified Medical Specialists*. New Providence, NJ: Marquis Who's Who Inc; 1997.
16. American Board of Medical Specialties. *The Official ABMS Directory of Board Certified Medical Specialists*. New Providence, NJ: Marquis Who's Who Inc; 1998.
17. American Board of Medical Specialties. *The Official ABMS Directory of Board Certified Medical Specialists*. New Providence, NJ: Marquis Who's Who Inc; 2000.
18. Schauer P, Ikramuddin S, Hamad G, Gourash W. The learning curve for laparoscopic Roux-en-Y gastic bypass is 100 cases. *Surg Endosc*. 2003;17:212-215.
19. Kligman MD, Thomas C, Saxe J. Effect of the learning curve on the early outcomes of laparoscopic Roux-en-Y gastric bypass. *Am Surg*. 2003;69:304-310.
20. Flum DR, Dellinger EP. Impact of gastric bypass operation on survival: a population-based analysis. *J Am Coll Surg*. 2004;199:543-551.
21. McMasters KM, Wong SL, Chao C, et al. Defining the optimal surgeon experience for breast cancer sentinel lymph node biopsy: a model for implementation of new surgical techniques. *Ann Surg*. 2001;234:292-299.
22. Bennett CL, Stryker SJ, Ferreira MR, Adams J, Beart RW. The learning curve for laparoscopic colorectal surgery: preliminary results from prospective analysis of 1,194 laparoscopic-assisted colectomies. *Arch Surg*. 1997;132:41-44.
23. Watson DI, Baigrie RJ, Jamieson GG. A learning curve for laparoscopic fundoplication: definable, avoidable, or a waste of time? *Ann Surg*. 1996;224:198-203.
24. Flum DR, Koepsell T, Heagerty P, Pellegrini CA. The nationwide frequency of major adverse outcomes in antireflux surgery and the role of surgeon experience, 1992-1997. *J Am Coll Surg*. 2002;195:611-618.
25. Moffat DA, Hardy DG, Grey PL, Baguley DM. The operative learning curve and its effect on facial nerve outcome in vestibular schwannoma surgery. *Am J Otol*. 1996;17:643-647.
26. Cochrane DD, Kestle JR. The influence of surgical operative experience on the duration of first ventriculoperitoneal shunt function and infection. *Pediatr Neurosurg*. 2003;38:295-301.

DISCUSSION

Myriam Curet, MD, Stanford, Calif: I want to congratulate you on an extremely elegant study evaluating competency and proficiency from a different prospective and a larger scale than has been done previously. Dr Prystowsky used an administrative state database to evaluate outcomes for over 120 000 alimentary tract operations in patients from 1996 to 1999. He used comorbid factors to stratify these into low- and high-complexity operations and then evaluated morbidity, mortality, and LOS by surgeon experience. He found that for high-complexity operations, surgeons who were less experienced had a mortality of 7% and a morbidity of 34%, compared to 5% mortality and 30% morbidity for surgeons who were more highly experienced. These results are quite troubling and I think will have significant impact on residency training programs, the ABS, the Residency Review Committee for Surgery, insurance companies, and obviously patients as well. The solutions are not obvious, and brings forth the issue of whether we are going to need longer length of training, more money to train surgeons better, and perhaps even limiting the practice of new surgeons after they are certified. I have several questions for the author.

First of all, when you stratified surgeon experience, you looked at 0 to 5 years, 6 to 20 years, 21 to 30 years, and over 31 years, with 60% of the surgeons falling in the 6- to 20-year range. I wondered why you chose this stratification and whether you initially did data analysis on the 5- to 10-year range. There are a lot of data in the education literature suggesting that it takes 10 years to become proficient at anything that one does and perhaps looking at the 6- to 10-year range would have been helpful in looking at these data.

Second, there are other factors that could have biased these results. For example, experienced surgeons will often have the same OR [operating room] team so that the circulating nurse, the technician, and the anesthesiologist know what the surgeon wants and how he or she wants it done and, therefore, they can be more efficient. Likewise, the experienced surgeon is more likely to have a more experienced PA [physician assistant] or perhaps a senior resident assisting them while the less experienced surgeon is more likely to be operating with someone who is also less experienced.

And finally, the experienced surgeon is more likely to have block time starting at 7:30 rather than having to wait until late at night for OR time to be available. I wondered whether you could comment on these factors.

You have also suggested in your "Comment" that the learning curve is a significant part of these data and that perhaps more experience is necessary for the younger surgeons to become more proficient. I think there is also the learning curve associated with being the senior person in the room, and these younger surgeons, until recently, have always had somebody who is more experienced in the room with them who knows how to get exposure, knows how to use assistants, and may have some tricks in their book for how to deal with a difficult situation. Perhaps one thing residency training programs could do would be to allow the residents to be the senior surgeon in the room and try to struggle through some of these factors before they get out into practice.

You have also suggested that perhaps these new surgeons need more supervision once they get out of training. Many hospitals actually do require that new surgeons are supervised for a specific period of time or a certain number of cases before allowing them independent privileges. Do you know whether any of these less experienced surgeons actually were supervised during their first few cases, because this would clearly impact those results?

Finally, I think that these data really beg the question of tracking and early specialization. If, for example, a third-year resident knows that he or she is going to go into cardiac surgery, those chief resident cases that he or she would normally do could now be given to the resident who is going to go into gastrointestinal surgery and, thus, perhaps they could get more experience and more cases during residency prior to starting out in practice. I think that this is the single most important issue that needs to be addressed when discussing these data.

Dr Prystowsky: The first question related to the length of time and choosing the designation of 0 to 5 years since ABS certification. I did try various formulae. The consideration that I had was to determine that period of time in which residency would have the greatest effect, hence, the 0 to 5 years. The other designations were more for convenience. Undoubtedly, there are a variety of variables that influence outcomes and you mention several of them in particular. No doubt, and this is the disadvantage as we all know of a retrospective review, that we cannot easily sift out the multitude of variables that can affect outcomes and that is a point well-taken.

With regard to supervision, I think the point is that in this age of the 80-hour workweek, as skill laboratories become prevalent and as we, I think, generally need to be as thoughtful as possible about the education of our residents, there are data that we can use to refine their experience to determine what they will be doing in the future and hopefully tailor some of their education within residency to those needs. That does not mean that a resident should not be doing high-complexity procedures to round out their education even if they are not doing them in the future, but I do think it requires us to be even more thoughtful than we have in the past about exactly what they do accomplish in residency given what they are going to be doing in practice subsequently. My bias is that as residents leave our programs, typically "out of sight, out of mind." We shake their hand, kiss them good-bye, and really do not know exactly how they perform subsequently, and there are ways in which we should be able to do that. Early specialization absolutely falls into this concept of trying to be as thoughtful as possible about preparing each individual for the type of practice they are going to have in the future.

David Rothenberger, MD, Minneapolis, Minn: Along the lines of your last comment and in light of President Michelassi's talk yesterday, did you have in your database the information about whether these young surgeons were in subspecialties? Had they focused themselves? Had they done specialty fellowships?

Dr Prystowsky: No, we did not look at, in this particular study, fellowship training and the like that the younger surgeons, or even the older surgeons, might have accomplished and how that would have affected their practice. I did perform a subset analysis for segmental colectomy looking at a number of variables, including colorectal fellowship, and for that particular procedure in this group of patients, colorectal fellowship did not have a favorable influence on outcomes. That is somewhat in contradistinction to literature related to more complex rectal procedures where it does have a favorable influence and one might say, well that makes sense since residents are getting to do a lot of colectomies during residency so they are prepared to do those cases, but that is the only subset analysis that would come close to answering your question.

Raymond Joehl, MD, Maywood, Ill: I am concerned that you have not adequately considered the impact of comorbid illnesses on mortality and morbidity. Adjusting raw mortality and morbidity rates is essential when comparing competencies of surgeons or groups of surgeons. Perhaps the best example of risk adjustment has been done at the VA [Veterans Affairs] National Surgical Quality Improvement Program, which the American College of Surgeons has embraced and is implementing in the private sector. I have one question. Is there a way to adjust for risk factors on surgical morbidity and mortality using the administrative database in this study?

Dr Prystowsky: It is very difficult in this kind of administrative database review to risk stratify patients. The attempt here was to use associated diagnoses to develop a severity of illness index. This was actually a fairly good predictor for death. But I take your general point and I think with the NSQIP [National Surgical Quality Improvement Program] directive from the VA now going into the private sector, that maybe 5, 10 years from now we will be able to get good data in a prospective fashion from a wide variety of hospitals that, in turn, can be used to look at the level of surgeon performance. I take your point, but the severity of illness index worked reasonably well in trying to stratify patients.

Tyler Hughes, MD, McPherson, Kan: Aviation and surgery are frequently compared. The first 100 hours after your pilot's license is known as the killing zone. Certainly a negative feedback on competence. One does not see that sort of killing zone effect in corporate airlines where there is a crew concept and a senior pilot flying with the more junior pilot. Does your data look at, what I have seen as a trend, that surgeons more and more are operating alone or with PAs? If these younger surgeons were operating with more experienced surgeons acting as assistants, would there be an equalization of the morbidity and mortality rate?

Dr Prystowsky: I think this is a good question. These data do not permit that sort of an analysis, so I do not really know, nor was this study specifically designed to look at learning curve per se. But I think it does raise the question; that is, if one accepts that learning curves exist for all procedures, and in this case, highlighted by the higher-complexity procedures, then it may be that a resident with clear guidelines would go out into practice and realize that they should have supervision for some additional number of cases. I am sure it is happening already in many circumstances, but I think perhaps not in as many as it should.

Bruce Gewertz, MD, Chicago, Ill: I would like to question 2 elements of this presentation.

One, is "competence" really the right word? Competence would not be the word that I would use to describe a slight difference in very sophisticated statistical analysis. I think the danger is political because when our patients and friends in the community hear we are suggesting that some physicians are not competent, it raises a whole bunch of issues that the term "performance difference" may not raise.

Finally, I wonder if you could comment on the statistical power of the difference that you showed? We are all aware of the fact that something can be different but not important and I wonder if you somehow quantitate the difference in terms of number of deaths or morbidity in this large group. In other words, does it count for 2 or 3 across the state in a year, or are we talking about hundreds of patients who are adversely affected by this?

Dr Prystowsky: I am 46 now and I definitely am a different surgeon than I was when I was 36. No doubt there is, to some extent, no substitute for experience that we are all aware. I think, though, that we should not fool ourselves into thinking that that can account for all the differences, and it is worth taking a critical look as to whether we can somehow close that gap.

In terms of competence, respectfully, I disagree. I will grant that the nature of this study is such that we may want to be circumspect about how much stock we put in it. However, these young surgeons operated on close to 10 000 people, so we have to be concerned about the difference in outcomes and look for ways to readdress the gap. In terms of the number of patients and exactly how clinically significant it is, I could probably do that analysis but I would have to get back to you next week.

Leigh Anne Neumayer, MD, Salt Lake City, Utah: This is a great paper and confirms all of my biases, but it also confirms other studies, for instance, in cardiac surgery looking at surgeons who are in their first 5 years and then, of course, the one that I am most familiar with, which is a prospective study, our big VA hernia trial. We saw a significant learning curve in the laparoscopic group needing greater than 250 cases in order to have a recurrence rate of 5%, which was equivalent to the open hernia repair group. As surgeons, we need to understand that these data are real. We need to figure out the nuances of these procedures in order to teach efficiently so we can truncate that learning curve. Because of the lack of resources and limited case volume, there are not that many cases out there and because it is not fair to the patients to have such long learning curves, we need to pay greater attention to the actual procedures. The statistician who worked on our study and was at the presentation in April at the Spring Meeting of the American College of Surgeons suggested afterwards that as surgeons maybe what we need to do is to have "masters" classes. Anybody that is a musician will know that a masters class is when the master comes in and watches you play. After we have graduated residency, it is very rare that an expert surgeon comes in and actually watches us operate.

Harold Gaskill, MD, San Antonio, Tex: It is nice to see papers with tens of thousands of cases, and I think a lot can be learned from them. If I were operating a residency training program in Illinois, I would be concerned that everybody else's residents were skewing the data from my residents. I am wondering if you were able to compare outcomes from residents at different kinds of programs (eg, university programs based at a tertiary care center vs private practice-based programs)?

Dr Prystowsky: I did not do that for all 120 000 patients here. I did do that for a subset of about 18 000 patients who underwent a segmental colon resection, and comparing type of residency program of the surgeon, specifically university-based vs a non–university-based program, found no difference in outcomes when you used that as a single variable.

Amilu Rothhammer, MD, Colorado Springs, Colo: There is one other variable that I think is pretty important. We brought up the difference between the rural surgeon and the community generalist. Also, I think as these people come up into practice, whether they come into a solo practice by themselves or with a group practice makes a great deal of difference on their complication rate.

Dr Prystowsky: I agree.

Sherry Wren, MD, Palo Alto, Calif: You reported a morbidity of 34%, but when you look at the length of stay, it is basically equal across those groups. What's happening there?

Dr Prystowsky: I noticed that as well, and it is very difficult to explain. The only trend that seemed to be present, at least in low-complexity procedures, is that when it comes to LOS, the newer surgeons tend to get their patients out faster. I am assuming that that might be a bias of older more experienced surgeons who might hang onto their patients a little bit longer and somehow that balances out. But I think it is a very good question, and I do not have a good answer for it.

Kenric Murayama, MD, Honolulu, Hawaii: You stratified based on number of years out of training and I wondered if there was a way in your database to look at actual surgeon experience and cross-reference that to years in practice since some surgeons who are out 20 years have not done many advanced cases or do 1 or 2 a year. If you could break down that group of surgeons, it would be interesting to see if the complication rates were higher for surgeons who did comparatively fewer advanced cases per year. Can this be done?

Dr Prystowsky: It would be nice, and I did not do it, to look at cumulative experience and there are statistical methods to do that. The only thing that I did that partially answers your question is that if a surgeon during this study period had done less than 10 alimentary tract operations for whatever reasons, I excluded their patients completely from this analysis. So I used a cutoff of 10 to avoid the surgeon who may do a rare case so that the data would not be unfairly biased because of that. But in terms of the cumulative experience over time, that is a very good question. I did not do it.

Fabrizio Michelassi, MD, New York, NY: I was wondering whether the same data applied in academic vs nonacademic centers. Also, it is not clear to me how you defined low-complexity vs high-complexity procedures. Finally, I wonder whether you were able to stratify mortality and morbidity for high-complexity procedures based on surgeon's age.

Dr Prystowsky: The definition of low and high complexity was purely based on appendectomy and cholecystectomy being defined as low-complexity procedures and everything else being defined as high complexity. I did not look at hospital variables, either the type of hospital, academic or tertiary care vs other types, or hospital volume that is a useful variable in outcome studies. Again, this begs for a second-order analysis of the surgeons, their training, cumulative experience, fellowship training, and the like that I did not do.