ARCHIVOS DE BRONCONEUMOLOGIA

Original Article

Evaluation of the Interobserver Variability in the Systematic Classification of Operative Morbidity in Lung Resection☆

Gonzalo Varela,* Nuria M. Novoa

Servicio de Cirugía Torácica, Hospital Universitario de Salamanca, Salamanca, Spain

A B S T R A C T

Objective: To quantify the interobserver variability in establishing a systematic classification for the operative morbidity of lung resection.
Methods: Morbidity was classified retrospectively in a series of 499 prospective registries of patients who underwent major lung resection (458 lobectomies and 51 pneumonectomies). The systematic classification proposed by Seely et al. in 2010 was used. Each one of the authors independently classified the complications and the weighted kappa statistic was calculated.
Results and comments: The kappa index was 0.79. Although the value is high, it introduces a systematic bias in the classification of patient morbidity that indicates the need to very carefully evaluate the data entered into the multi-institutional registers in order to be able to obtain valid conclusions.

© 2011 SEPAR. Published by Elsevier España, S.L. All rights reserved.

Evaluación de la variabilidad interobservador en la clasificación sistemática de la morbilidad operatoria en resección pulmonar

R E S U M E N

Objetivo: Cuantificar la variabilidad interobservador al establecer una clasificación sistemática de la morbilidad operatoria de la resección pulmonar.
Método: Se ha clasificado la morbilidad de forma retrospectiva en una serie de 499 registros prospectivos de pacientes sometidos a resección pulmonar mayor (458 lobectomías y 51 neumonectomías). Se utilizó la clasificación sistemática propuesta por Seely y col en 2010. Cada uno de los autores clasificó de forma independiente las complicaciones y se calculó el estadístico kappa ponderado.
Resultados y comentarios: El índice kappa fue de 0.79. Aunque el valor es alto, introduce un sesgo sistemático en la clasificación de la morbilidad de los pacientes que indica la necesidad de valorar muy cuidadosamente los datos introducidos en los registros multiinstitucionales para poder obtener conclusiones válidas.

© 2011 SEPAR. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Introduction

For the evaluation of the quality of surgery, it is essential to understand its possible adverse effects, fundamentally operative morbidity and mortality. A previous study done in Spain on patients who underwent lung resection (in which administrative databases were used)[1] demonstrated that there are notable differences between centers when comparing hospital mortality, both gross as well as risk-adjusted. Mortality is a very robust variable in administrative databases because patient deaths are always recorded. When post-surgery morbidity was analyzed, however, the results were in somewhat disconcerting as we found that the centers with greatest mortality did not register the greatest morbidity, which is what was expected. This finding, which could be due to variations in local clinical practices and the quality of the clinical registries, makes one consider that perhaps the analysis of the morbidity of large patient series incorporated in multicenter databases could be subjected to a systematic bias that could invalidate the conclusions.

Recently, Seely et al.[2] have proposed the application of a systematic classification of operative morbidity into five standardized

Table 1
Systematic Classification of the Complications in Thoracic Interventions.

Degree	Definition
Minor complication	
I	Pharmacological treatment is not necessary, nor is any other
II	Requires pharmacological treatment or minor intervention
Major complication	
III	Requires a surgical, radiological or endoscopic intervention or several treatments
IIIa	The intervention does not require general anesthesia
IIIb	The intervention requires general anesthesia
IV	Needs treatment in the intensive care unit and life support
IVa	Dysfunction of one organ
IVb	Dysfunction of multiple organs
V	The complication leads to patient death

Modified by Seely et al.[2]

groups (Table 1). They conclude that the use of this type of classification facilitates the objective comparison between surgical procedures and series of patients, as well as between surgeons and different surgical groups. However, although the authors refer to this subject in their publication, the interobserver variability when classifying post-surgical complications is unknown, which is fundamental before adopting this type of classification for the analysis of large clinical databases.

In this study, our intention is to quantify the agreement between two independent observers when classifying postoperative complications in lung resection.

Methods

Ours is a retrospective analysis of the registers of all the patients who had undergone anatomical lung resection (lobectomy or pneumonectomy) in our center between January 2007 and December 2010. The registers were obtained from a prospective database that contains the data of all the patients who have undergone surgery from January 1994 to date. The quality of the data is ensured by two quality controls performed by a data manager: the first is when the release report is emitted, and the second is done before the clinical documentation is sent to the general hospital archives after having incorporated any pending histologic studies, etc. The postoperative complications are coded in accordance with a document defining each type of post-surgical complication.

For this study, we utilized the information contained in the database that include the following variables: medical file reference number, main surgical procedure, mortality within 30 days of the procedure, postoperative complication (yes/no), type of complication (up to 4 complication codes per patient) and, in addition, the need for repeated surgery and the death of the patient within 30 days following the operation. A manual review revealed any incongruencies or missing data seen in the initial list and established the definitive list, which had no missing data or incongruencies.

Once the data had been reviewed, a list was prepared of cases with their complications. Two authors independently classified the complications into 5 levels, in accordance with the definitions presented in the article by Seely et al.[2] Finally, the two gradations were united into one single archive and the Cohen's weighted kappa statistic was calculated. In order to quantify the importance of the disagreement, the default values provided by the statistical program used (Stata 10.1) were accepted.

Table 2
Complications Registered in the Series of Cases.

Type of Complication	Complication Registered (n of Cases)					
	A	B	C	D	E	F
Cerebrovascular accident	1					
Cardiac arrhythmia	18	4	1			
Atelectasis	5	1	3			
Pulmonary edema	1					
Pulmonary embolism	1					
Pleural empyema	4	3	1			
Bronchial fistula	2	1	2			
Air leak >5 days	45	3	1			
Wound hematoma	1					
Hemorrhage	2	1				
Hemothorax	8	1				
Diaphragmatic hernia		1				
Paralytic ileus	2					
Wound infection	6	4		2		
Urinary infection	2					
Heart failure	7	1	2			
Kidney failure	2		1	1		
Peripheral	3	1				
Need for oxygen therapy upon discharge	5	4	1			
Nosocomial pneumonia	9	5	1			
Pneumothorax	7	2				
Cardiac arrest	1					
Pericarditis	1					
Reaction to medication	1					
Urine retention	1					
Sepsis	1					
Post-op mechanical ventilation	1					
Major re-intervention					18	
Death						5

Table 3
Classifications Assigned by the Two Observers.

Observer 2	Observer 1							
	I	II	IIIa	IIIb	IVa	IVb	V	Total
I	40	5	0	0	0	0	0	45
II	1	43	4	0	0	0	0	48
IIIa	0	8	10	0	0	1	0	19
IIIb	0	1	1	12	0	0	0	14
IVa	0	1	1	1	2	0	0	5
IVb	0	2	0	2	0	0	0	4
V	0	0	0	0	0	0	5	5
Total	41	60	16	15	2	1	5	140

Kappa 0.79 (standard deviation, 0.05; *P* < 0.0001).

Results

A series of 499 patients were obtained (41 pneumonectomies and 458 lobectomies). The 30-day mortality was 1% (5 patients, 2 post-lobectomy and 3 post-pneumonectomy). One or more complications were registered in 140 cases (28%). The list of complications used by the evaluators is shown in Table 2.

The classification of the complications into levels, as carried out by the two observers, is shown in Table 3. The valor weighted kappa statistic is 0.79 (standard deviation, 0.05; *P* < 0.0001).

Discussion

The analysis of large clinical databases is currently a necessary medium for the improvement of quality health care. In a short time, it provides robust data based on large patient populations that serve as a standard for comparing the data of the institution itself and for introducing any corrective measures necessary in cases when important deviations are observed.[3,4] Although it is true that the quality of surgical work should not be solely based on the results (mortality and morbidity), until now there are very few initiatives

published on the use of other variables, such as the appropriateness of the design of clinical processes to the best scientific evidence available. In thoracic surgery, it has been proposed to construct an index that gathers the results and designs of processes in order to evaluate their quality,[5] but its practical application is still being developed.[6]

The complication severity classification proposed by Seely et al.[2] seems, *a priori*, to be an adequate method for standardizing the evaluation of the results of operative morbidity, giving the comparison between centers greater validity. In addition, it introduces an innovative concept as it considers the resources used to treat complications.

The calculation of the kappa index for inter-observer agreement is a widely used method in clinical epidemiology when quantifying the variability between various observers of clinical events. In addition, it is a much more representative parameter than the simple agreement rate, as it is very much influenced by chance.[7] When the order of the values that are obtained in the observation is important, as is the case of this study, it is recommendable to use the weighted kappa index, since it is not equally evaluable if there is a discrepancy classifying the complication as degree 1 or degree 2, or of the discrepancy is between grades 1 and 5. Although the use of the kappa statistic has been criticized for its dependence on the prevalence of the event studied and for the subjectivity of the weighting of the discrepancies,[8] it is considered that a weighted kappa index above 0.8 indicates an excellent degree of inter-observer agreement.[7] In our case, the kappa index obtained of 0.79 is worthy of mention.

Although without a doubt this value demonstrates that there is a high agreement between two observers when classifying the severity of the registered complications, it should be kept in mind that the two observers participate in daily discussions of the same cases, and that in the center there is a previous, agreed-upon definition of the complications. In this context, the expected inter-observer agreement should probably be 100%. It can be easily supposed that, when working with a multi-institutional database, the degree of agreement when defining and classifying the complications should be much less or, in other words, the effort required for the validation of the data when analyzing morbidity would be much greater.

The study that we have presented demonstrates that, even between surgeons of the same center, there is a certain degree of discrepancy in the classification of the post-surgical complications. Therefore, before adopting a standardized system for classifying morbidity at the multi-institutional level, it is necessary to study the variability and introduce the necessary corrective measures in order to construct risk indices in thoracic interventions. These measures would include the precise and unequivocal definition for each type of complication and a generic classification of the treatments required to resolve them. Only by guaranteeing the quality of the data entered in the clinical databases, valid conclusions can be obtained.[9]

References

1. Varela G, Molins L, Astudillo J, Borro JM, Canalís E, Freixinet J, et al. Experiencia piloto de benchmarking en Cirugía Torácica: comparación de la casuística e indicadores de calidad en resección pulmonar. Arch Bronconeumol. 2006;42:267–72.
2. Seely AJE, Ivanovich J, Threader J, Al-Hussaini A, Al-Shehab D, Ramsay T, et al. Systematic classification of morbidity and mortality alter thoracic surgery. Ann Thorac Surg. 2010;90:936–42.
3. Brunelli A, Varela G, Berrisford R, Rocco G. Audit, quality control and performance in thoracic surgery: an European perspective. Thorac Surg Clin. 2007;17:387–93.
4. Brunelli A, Varela G, Van Schil P, Salati M, Novoa N, Hendriks JM, et al. on behalf of the ESTS Audit and Clinical Excellence Committee. Multicentric analysis of performance after major lung resections by using the European Society Objective Score (ESOS). Eur J Cardiothorac Surg. 2008;33:282–8.
5. Brunelli A, Berrisford RG, Rocco G, Varela G, on behalf of the European Society of Thoracic Surgeons Database Committee. The European Thoracic Database project: composite performance score to measure quality of care after major lung resection. Eur J Cardiothorac Surg. 2009;35:769–74.
6. Brunelli A, Rocco G, Van Raemdonck D, Varela G, Dahan M. Lessons learned from the European thoracic surgery database: the composite performance score. Eur J Surg Oncol. 2010;36 Suppl. 1:S93–9.
7. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation and sample size requirements. Phys Ther. 2005;85:267–8.
8. Thompson WD, Walters SD. A reappraisal of the kappa coefficient. J Clin Epidemiol. 1998;41:949–58.
9. Pass HI, Medical registries. Continued attempts for robust quality of data. J Thorac Oncol. 2010;5:S198–9.