

## TRABAJOS CIENTÍFICOS

# Evaluación de resultados quirúrgicos. Validación inicial del escore POSSUM\* Use of the physiological an operative severity score for the enumeration on mortality and morbidity (possum) score in a surgical service

Drs. VÍCTOR MOLINA G<sup>1</sup>, LUIS GARRIDO O<sup>1</sup>, CARLOS MANTEROLA D<sup>1,2</sup>

<sup>1</sup>Departamento de Cirugía, Facultad de Medicina, Universidad de La Frontera, Chile.

<sup>2</sup>Capacitación, Investigación y Gestión para la Salud Basada en Evidencia (CIGES)

### RESUMEN

*Introducción:* Los procesos de evaluación son una actividad fundamental dentro de los procesos del área de la salud. Nos permiten pesquisar errores y corregirlos, y detectar conductas acertadas y reforzarlas. Una forma de evaluar esto es mediante la aplicación de sistemas de puntuación ad-hoc, como el escore POSSUM. El objetivo de este trabajo es validar el escore POSSUM en pacientes quirúrgicos atendidos en la Unidad de Urgencia y Servicio de Cirugía del Hospital Regional de Temuco. *Material y método:* Estudio de corte transversal. Se realizó la traducción del instrumento original, y posteriormente se aplicó éste a una muestra no-probabilística de pacientes provenientes de un muestreo aleatorio entre aquellos sujetos mayores de 18 años, sin restricción de género, que fueron intervenidos quirúrgicamente en los Servicios de Cirugía y Urgencia del Hospital Regional de Temuco, entre diciembre de 2002 y octubre de 2003. A los pacientes se les aplicó el escore en forma independiente por dos observadores con similar grado de entrenamiento, se evaluó la comprensión de los ítems, el tiempo para la aplicación del instrumento, la necesidad de entrenamiento y el grado de acuerdo ínter observador. Posteriormente, se evaluó la consistencia interna de la escala mediante la aplicación del alfa de Cronbach. *Resultados:* No existieron problemas de interpretación en la traducción del instrumento. En la comprensión de los ítems no hubo problemas significativos. El tiempo para aplicar la escala en promedio fue de 7 minutos por paciente. Encontramos un acuerdo en más del 83% de las mediciones en todas las variables fisiológicas, excepto en la interpretación del electrocardiograma en donde se produjo acuerdo sólo en el 40,5%. En la parte quirúrgica del escore se verificaron acuerdos de 85,7% para "presencia de cáncer", 75% para "contaminación del sito operatorio", 87,5% para "sangrado intraoperatorio", y 50% para "magnitud de la cirugía". La evaluación de la homogeneidad de la escala permitió verificar un alfa de Cronbach de sólo 0,52 para las variables fisiológicas, y de 0,47 para las variables quirúrgicas. *Conclusión:* Los resultados obtenidos deben reforzar el concepto de validación de todo escore foráneo independiente del rendimiento reportado en su lugar de origen. Sin duda nuestra realidad puede ser muy diferente a aquella en la que se generó el instrumento.

PALABRAS CLAVES: **POSSUM, escores, escalas, pronóstico, evaluación.**

\*Recibido el 20 de diciembre de 2004 y aceptado para publicación el 25 de abril de 2005.

E-mail: cmantero@ufro.cl

### SUMMARY

Background: Assessment of the overall results of surgical services is of utmost importance. The POSSUM score allows a global assessment of surgical results, independent of the complexity of the surgical procedures performed. Aim: To report the result of the use of the POSSUM score in a surgical service. Material and methods: Cross sectional application of POSSUM score in a random sample of patients over 18 years of age in a Surgical Service of a public hospital, between December 2002 and October 2003. Each patient was assessed by two observers, that were unaware of the results of each other. The time required to apply the score and the concordance between observers was measured. The internal consistency of the scale was assessed using Cronbach's alpha. Results: The instrument was well translated and all items were correctly understood. It was applied to 180 patients (mean age 49 years, 90 males). The mean time required to apply the score was seven minutes. Concordance for the measurement of physiological variables was 83%, except for the interpretation of EKG, whose concordance was 50%. There was a 86% concordance for the item "presence of cancer", 75% for "contamination of operative site", 86% for "intraoperative bleeding" and 50% for "magnitude of the surgical procedure". Internal consistency was 0.52 for physiological variables and 0.47 for surgical variables. Conclusions: This score can be applied in surgical patients in Chile but some variables have a low concordance between observers.

KEY WORDS: **POSSUM, scores, surgical results.**

### INTRODUCCIÓN

Los procesos de evaluación son una actividad fundamental dentro del desarrollo del conocimiento. Estos nos permiten, por una parte pesquisar errores y corregirlos, y por otra, detectar conductas acertadas y reforzarlas.

En el área de la salud este concepto cobra gran importancia, sin embargo, es el escenario donde más difícil resulta realizar evaluaciones objetivas. Por ejemplo, en el ámbito de la cirugía y disciplinas afines, los resultados son medidos clásicamente en base a morbilidad y mortalidad, lo cual ciertamente no refleja la calidad del acto quirúrgico.

Por razones como la anteriormente expuesta, se han desarrollado sistemas de puntuación o escores, cuyo propósito es permitir una valoración objetiva de los procesos en el ámbito sanitario. El sistema de escore ideal debe ser capaz de predecir morbilidad y mortalidad, ser rápido y fácil de usar, y debe ser aplicable en un amplio espectro de situaciones.

Uno de los campos en que se han desarrollado estos sistemas de puntuación es en lo que a riesgo se refiere, escenario de gran relevancia en la práctica cotidiana de la cirugía.

Uno de estos escores fue desarrollado en Inglaterra durante los años 90, se denominó sistema POSSUM (Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity), y ha sido utilizado para evaluar calidad de resultados en cirugía, en términos de estimar la probabilidad de aparición de morbilidad y mortalidad para grupos de pacientes según el estado fisiológico preoperatorio y la magnitud de la injuria quirúrgica<sup>1</sup>. De esta forma, el escore POSSUM ha

hecho posible la comparación de resultados entre cirugías de distinto nivel de complejidad.

El escore consta de 2 dominios, uno que evalúa el estado fisiológico del paciente, y el otro que valora el área quirúrgica. La evaluación del estado fisiológico se compone a su vez de doce ítems. Estos, incluyen síntomas y signos clínicos, y algunos exámenes simples (bioquímicos, hematológicos, electrocardiograma). Cada uno de estos ítems constituye una variable dividida en 4 categorías con un incremento exponencial del puntaje en la medida que se aleje de la normalidad (1, 2, 4 y 8 puntos). Es así como se puede obtener un puntaje mínimo de 12 y uno máximo de 88 puntos. Por su parte, la evaluación de la parte quirúrgica se compone de seis ítems, los que reflejan la gravedad de la injuria operatoria. Del mismo modo, estas variables se dividen en cuatro categorías que reciben puntuaciones de 1, 2, 4 y 8.

El puntaje total del escore (la suma del puntaje fisiológico más el quirúrgico), no refleja el porcentaje de riesgo. La estimación del riesgo se obtiene aplicando dos ecuaciones, una para morbilidad y otra para mortalidad. Los resultados representan una estimación de la probabilidad de mortalidad y morbilidad a 30 días<sup>1</sup>.

El principal valor de este escore, además de su simpleza, es el riguroso método científico a través del cual fue desarrollado<sup>1-3</sup>, y parece reunir las características descritas para un buen escore. Sin embargo, se requiere de una adecuada validación y estudios de confiabilidad antes de poder proponer su uso en nuestro medio.

Por ende, nos planteamos la siguiente pregunta de investigación: ¿es aplicable el escore

POSSUM para la evaluación de resultados quirúrgicos en el Hospital Regional de Temuco? Por ello, el objetivo de este trabajo es validar el score POSSUM en los pacientes quirúrgicos atendidos en la Unidad de Urgencia y Servicio de Cirugía del Hospital Regional de Temuco. Para lo cual, en esta primera etapa decidimos trabajar con los siguientes objetivos específicos: Traducir el instrumento original, evaluar la aplicabilidad del instrumento y evaluar la confiabilidad del instrumento; dejando para la siguiente etapa del estudio la aplicación de pruebas de validez al score de POSSUM.

### MATERIAL Y MÉTODO

*Diseño:* Estudio de corte transversal.

*Población:* Se reclutaron pacientes provenientes de un muestreo aleatorio entre aquellos sujetos mayores de 18 años, sin restricción de género, que fueron intervenidos quirúrgicamente en los Servicios de Cirugía y Urgencia del Hospital Regional de Temuco, durante el periodo comprendido entre diciembre de 2002 y octubre de 2003. A todos ellos, o su familiar más cercano, se le solicitó el consentimiento informado para la participación en el estudio.

*Estimación de la muestra:* La validación de toda escala requiere de una muestra mínima de 10 pacientes por cada uno de los ítems que la componen. La escala a evaluar consta de 18 ítems, por lo tanto, nuestro tamaño de muestra fue de 180 pacientes.

*Plan del análisis:*

1. Selección de un instrumento de medición. Se buscó dirigidamente en la literatura el mejor score disponible y que presentara las siguientes características: que considerara variables susceptibles de medir en nuestro medio, rápido y fácil de aplicar, validado en su lugar de origen, y que fuera aplicable en un amplio espectro quirúrgico.

2. Traducción del instrumento de medición. Se realizó una traducción inglés-español, español-inglés e inglés-español por tres observadores en forma independiente.

3. Realización de prueba piloto. Con una muestra por conveniencia, de 180 pacientes, que fueron encuestados en forma separada por dos observadores con similar grado de entrenamiento, se evaluó la comprensión de los ítems, el tiempo para la aplicación del instrumento, la necesidad de entrenamiento y el grado de acuerdo ínter observador.

4. Pruebas de confiabilidad. Se evaluó la consistencia interna de la escala. Para ello, se utilizó el alfa de Cronbach como medida de homogeneidad.

### RESULTADOS

1. *Caracterización de la población estudiada.* Durante el periodo de estudio se reclutaron 180 pacientes intervenidos en el servicio de Cirugía y Urgencias Adultos del Hospital Regional Temuco. El promedio de edad fue de 49 años, siendo el 50% de género masculino. El 63% de los casos correspondieron a cirugías realizadas en el Servicio de Urgencias. En el 48% de las intervenciones el cirujano fue un residente de cirugía general. Se registraron 6 muertes en el grupo y un 26% de los pacientes presentaron complicaciones. Los principales diagnósticos fueron apendicitis aguda (25% de los casos) y colecistitis aguda (13% de los casos).

2. *Traducción del instrumento original:* Por tratarse de variables con categorías en su mayoría numéricas, no existieron problemas de interpretación en la traducción del instrumento. Hubo acuerdo entre los observadores en la interpretación y traducción de aquellas variables descriptivas como el estado cardiovascular y respiratorio, y el electrocardiograma.

3. *Comprensión de los ítems.* la gran mayoría de las variables se refieren a resultados de pruebas de laboratorio por lo que no presentaron ninguna dificultad de interpretación. En la parte fisiológica las variables "estado cardiovascular y estado respiratorio" presentaron algunos problemas de interpretación de las categorías que las componen. Se presentaron diferencias de interpretación entre la normalidad y la categoría inmediatamente superior. En las categorías extremas la clínica es más evidente por lo que no presentaron mayor dificultad. En la parte quirúrgica no hubo problemas de interpretación de las variables y sus categorías.

4. *Tiempo de aplicación de la escala:* El tiempo para aplicar la escala en promedio fue de 7 minutos por paciente con un mínimo de 3 y un máximo de 15 minutos.

5. *Entrenamiento de los encuestadores.* Ambos encuestadores estimaron que las instrucciones iniciales para aplicar el instrumento fueron suficientes y no requirieron de entrenamiento especial.

6. *Confiabilidad ínter observador.* Se analizó la concordancia en aquellas variables clínicas que podrían estar sujetas a variaciones según la interpretación de cada observador. En la parte fisiológica se evaluaron las variables del estado cardiovascular y respiratorio, escala de coma de Glasgow y la interpretación de electrocardiograma. En la parte quirúrgica se evaluaron la magnitud de la cirugía, la hemorragia intra operatoria, la contaminación en el sitio operatorio, la existencia de cáncer y la oportu-

**Tabla 1**  
**ESCORE POSSUM. TRADUCCIÓN DE LOS ASPECTOS FISIOLÓGICOS**

Variables	Puntos			
	1	2	4	8
Edad (años)	< 61	61-70	> 70	
Signos cardíacos	Normal	Uso de diuréticos, digoxina, antianginosos o terapia antihipertensiva	Edema periférico, terapia con warfarina	Yugular ingurgitada
Radiografía de tórax			Cardiomegalia límite	Cardiomegalia
Signos respiratorios	Normal	Disnea de esfuerzo	Disnea de pequeños esfuerzos	
Radiografía de tórax		EPOC leve	EPOC moderada	Disnea reposo (frecuencia de 30/min) Fibrosis o consolidación
Presión arterial sistólica (mmHg)	110-130	131-170 ó 100-109	> 170 ó 90-99	< 90
Pulso (latidos x min.)	50-80	81-100 ó 40-49	101-120	> 121 ó < 40
Escala de coma Glasgow (puntos)	15	12-14	9-11	< 9
Urea (mg/dl)	< 50	50-75	76-100	> 100
Sodio (meq/l)	> 135	131-135	126-130	< 126
Potasio (meq/l)	3,5-5	3,2 - 3,4 ó 5,1 - 5,3	2,9-3,1 ó 5,4-5,9	<2,9 ó > 5,9
Hemoglobina (g/100 ml)	13-16	11,5-12,9 ó 16,1-17,0	10,0-11,4 ó 17,1-18,0	< 10 ó > 18
Recuento de leucocitos (xx)	4-10	10,1-20,0 ó 3,1-3,9	> 20 ó < 3,1	
Electrocardiograma	Normal		Fibrilación auricular (frecuencia de 60 a 90)	Cualquier otro ritmo anormal o >4 extrasístoles por min, ondas Q, alteraciones de la onda T o del segmento ST

nidad de la cirugía. No se encontró una diferencia significativa al analizar los puntajes asignados por ambos observadores tanto en la parte fisiológica como quirúrgica del score. Al analizar cada variable individualmente, encontramos un acuerdo en más del 83% de las mediciones en todas las variables fisiológicas, excepto en la interpretación del

electrocardiograma en donde se produjo acuerdo sólo en el 40,5% de las evaluaciones. En la parte quirúrgica del score se verificó un acuerdo de 85,7% para la variable "presencia de cáncer", de 75% para la variable "contaminación del sitio operatorio" y de 87,5% para la variable "sangrado intraoperatorio". Se debe aclarar que en el caso de

**Tabla 2**  
**ESCORE POSSUM. TRADUCCIÓN DE LOS ASPECTOS OPERATORIOS**

Variables	Puntos			
	1	2	4	8
Magnitud de la cirugía	Menor	Intermedia	Mayor	Mayor +
Procedimientos múltiples	1		2	> 2
Cantidad de sangre perdida (ml)	< 101	101-500	501-999	> 999
Contaminación peritoneal o del sitio operatorio	No	Menor (fluido seroso)	Pus localizado	Pus, sangre o contenido intestinal libre
Presencia de cáncer	No	Cáncer primario	Metástasis linfodiales	Metástasis a distancia
Oportunidad de la cirugía	Electiva		Urgencia: Cirugía en < de 24 h desde el ingreso (posible reanimación)	Urgencia inmediata: Cirugía en < de 2 h desde el ingreso

esta última variable, en la gran mayoría de las observaciones no se encuentra registrado en la descripción operatoria una estimación de ella, por lo que cada observador le asignó el puntaje mínimo. Para la variable "magnitud de la cirugía", hubo un acuerdo sólo de 50% en las mediciones.

7. *Consistencia interna.* La evaluación de la homogeneidad de la escala permitió verificar un alfa de Cronbach de 0,52 para las variables fisiológicas, y de 0,47 para las variables quirúrgicas. Esta última subía a 0,52 al eliminar la variable "presencia de cáncer". En los demás casos, no se producían variaciones significativas en las pruebas de homogeneidad al eliminar otras variables.

### DISCUSIÓN

Las auditorías profesionales tanto individuales como colectivas se han vuelto indispensables en la práctica quirúrgica actual<sup>2</sup>. En países desarrollados es común ver como los pacientes exigen conocer los resultados de cada cirujano antes de decidir a quien consultar. Chile no está ajeno a esta realidad y cada vez se hará más frecuente enfrentarnos a pacientes mejor informados y más demandantes de información acerca de nuestro desempeño.

Clásicamente los resultados en cirugía se han medidos en términos de morbilidad, mortalidad y estadía hospitalaria<sup>1</sup>. Sin embargo, establecer comparaciones de calidad usando solo estos elementos puede ser engañoso. No todos los casos presentan el mismo nivel de complejidad. Hospitales de urgencia y centros de derivación concentrarán los casos más complejos y, por lo tanto, su morbilidad, mortalidad y estadía hospitalaria serán mayores. Solo si las estimaciones de mortalidad y morbilidad incorporan un ajuste según el riesgo de cada paciente, o grupo de pacientes, será posible establecer comparaciones objetivas entre servicios quirúrgicos, hospitales y cirujanos<sup>1</sup>.

Los scores de riesgo buscan cuantificar el riesgo de cada paciente de morir o presentar morbilidad basados en la gravedad de su enfermedad.

El score POSSUM, diseñado en Gran Bretaña en los años 90', ha sido ampliamente usado en su país de origen para la realización de auditorías quirúrgicas. Al revisar la literatura aparece como el score más idóneo para este propósito. Parece tentador intentar su aplicación en nuestro medio, sin embargo, si actuamos con rigurosidad científica, esto no es posible sin una adecuada validación. Esto es lo que motivó nuestro estudio.

El estudio piloto demuestra que el score es sencillo y fácil de aplicar sin requerir un mayor

grado de entrenamiento e independiente del grado académico del encuestador. En este estudio los observadores fueron un interno de medicina de 7º año y un residente de cirugía general, logrando un porcentaje de acuerdo sobre un 83 % en la mayoría de las variables. La interpretación del electrocardiograma, que presentó un acuerdo ínter observador muy bajo (43,5% de acuerdo), exige que, al menos en esta variable, exista un mayor grado de entrenamiento definiendo claramente normalidad y anormalidad en sus diferentes grados.

A pesar de estar compuesto por variables simples, muchas de estas no son medidas de rutina en el preoperatorio. Entre estos destacan las mediciones de electrolitos plasmáticos y uremia. En la parte quirúrgica destaca la falta de registro del sangrado estimado durante la cirugía. Sin embargo, en aquellos casos en que hubo una hemorragia significativa, sí se encontró una estimación en la descripción operatoria. Parece lógico entonces, al menos en lo que respecta a esta variable, que se asigne el puntaje mínimo a aquellos casos sin registro, asumiendo una pérdida hemática poco significativa. Por otra parte, en las pruebas de homogeneidad, se pudo comprobar que el eliminar estas variables no producía mayor variación en el valor del alfa de Cronbach, por lo tanto, se podría asumir que son variables prescindibles.

En Chile no existe una categorización de las distintas cirugías según su grado de complejidad. Es por esto, que se decidió usar la categorización empleada en los trabajos originales para este fin. Esto puede ser discutible, y quizá se requiera previamente de un consenso de expertos para designar la complejidad de las distintas cirugías en nuestra realidad.

Las pruebas de consistencia interna estuvieron por debajo lo aceptable a pesar de haber usado un tamaño de muestra adecuado para la validación de esta escala (180 pacientes). Esto significa que las variables no representarían adecuadamente los fenómenos que pretenden medir (*estado fisiológico preoperatorio e impacto del acto quirúrgico*). Sin embargo, el resultado puede ser solo reflejo de la heterogeneidad de variables que representan a cada dominio. Por ejemplo, parece lógico que exista escasa relación entre variables como pulso y electrolitos plasmáticos. Nos parece que este resultado del alfa de Cronbach no invalida el uso de la escala y es solo reflejo de la diversidad de las variables que la componen.

A modo de conclusiones, se podría señalar que:

1. Los resultados demuestran que este score es aplicable en nuestro medio, sin embargo, si es-

peramos obtener un rendimiento similar al reportado en su país de origen, se requiere de rigurosidad en el registro de las distintas variables.

2. La falta de registro dificulta la aplicación no solo de este score en particular, si no que, frustra cualquier intento por lograr evaluaciones objetivas con cualquier sistema de medición. Esto no debe ser motivo para desechar un score, ni la solución debe ser simplificar aún más la escala.

3. Es necesario reevaluar aquellas variables de baja confiabilidad, lo cual está considerado en la segunda etapa de este estudio (aplicación de pruebas de validez).

4. Los resultados obtenidos deben reforzar además, el concepto de validación de todo score foráneo independiente del rendimiento reportado en su lugar de origen. Sin duda nuestra realidad puede ser muy diferente a la de aquellos y resulta preocupante que muchos de estos instrumentos de medición sean usados sin cuestionar siquiera su validez.

## REFERENCIAS

1. Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. *Br J Surg* 1991; 78: 356-60.
2. Neary WD, Heather BP, Earnshaw JJ. The Physiological and Operative Severity Score for enumeration of Mortality and Morbidity (POSSUM). *Br J Surg* 2003; 90: 157-65.
3. Copeland GP. The POSSUM System of Surgical Audit. *Arch Surg* 2002; 137: 15-19.
4. Sánchez Pedraza R, Gómez Restrepo C. Conceptos básicos sobre validación de escalas. *Rev Col Psiquiatría* 1998; 27: 121-30.
5. Streiner DL. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry* 1993; 38: 140-48.
6. Rosner B. *Fundamentals of Biostatistics*. 4th ed. Belmont: Duxbury Press 1994.
7. Feinstein A. *The theory and evaluation of sensibility. Clinimetrics*. Alvan Feinstein. Yale University Press, 1987.
8. Dawson-Saunders B, Traop GR. *Bioestadística Médica*. Mexico; Editorial Manual Moderno, 1993.