

Construcción y validación de un índice de gravedad para pacientes hospitalizados en áreas clínicas

R.E. Jiménez Paneque / J. Vázquez García / H. Fariñas Seijas

Hospital Clínico Quirúrgico «Hermanos Ameijeiras». Ciudad de La Habana. Cuba.

Correspondencia: Rosa E. Jiménez Paneque. Hospital C.Q. «Hermanos Ameijeiras». Sección de Investigaciones. San Lázaro 701, Ciudad La Habana 10300. Cuba

Recibido: 19 de julio de 1995

Aceptado: 4 de septiembre de 1996

(Development and validation of a severity of illness index for hospitalized patients in clinical areas)

Resumen

Las escalas de gravedad de la enfermedad para pacientes hospitalizados son utilizadas con frecuencia en la evaluación del desempeño de los servicios hospitalarios. El objetivo del trabajo fue desarrollar y validar un índice cuantitativo que midiera la gravedad de pacientes durante su etapa de hospitalización y que fuera sencillo de obtener, a partir de los datos de la historia clínica. La construcción incluyó: la selección de componentes y la búsqueda de ponderaciones. Para esta fase se consultó la literatura y a especialistas y 74 historias clínicas de pacientes dados de alta proveyeron la base empírica. Como resultado se obtuvo un índice con dos variantes, uno cuantitativo y otro cualitativo ordinal con cuatro niveles. La validación incluyó cuatro aspectos de la validez, y la confiabilidad interna. Dieciséis especialistas evaluaron validez de contenidos y presentación. Cien historias clínicas sirvieron de base para evaluar la validez por construcción, de criterio y la confiabilidad. La validez fue satisfactoria en la mayor parte de sus aspectos. El índice de confiabilidad fue de 0,95. La concordancia entre jueces arrojó un coeficiente kappa de 0,4 para el índice ordinal y los coeficientes de correlación mayores de 0,93 entre todos los pares de jueces. El indicado se considera listo para aplicar en este contexto aunque se hacen algunas sugerencias para mejoras en el futuro.

Palabras clave: Índice de gravedad de la enfermedad. Calidad de la atención médica. Análisis de regresión. Confiabilidad. Validez.

Summary

Severity of illness indexes for hospitalized patients are frequently used for the assessment of hospital performance. The purpose here was to develop and validate a quantitative index for clinical areas which could measure severity of illness during hospitalization period and would be easy to obtain from the information contained in a regular clinical chart. The construction included item selection and search for items weights. Experts and literature were consulted and 74 clinical records provided empirical information. The result was the proposal of an index with two alternatives: one quantitative and the other ordinal with four levels of severity. Validation included four aspects of validity, general reliability, interrater agreement and internal consistency. Sixteen specialized physicians assessed content and face validity. One hundred clinical records of discharged patients were used to assess criterion, construct validity and reliability. Results show satisfactory validity in almost all aspects. The reliability coefficient was 0.95, Kappa coefficient was 0.4 for the ordinal index and correlation coefficients over 0.93 for the quantitative one. The index is ready for current applications in this context although some suggestions for future improvements were also included.

Key words: Severity of illness index. Quality of care. Case-mix. Regression analysis. Reliability. Validity.

Introducción

La determinación cuantitativa o semicuantitativa de la gravedad de la enfermedad que afecta a un paciente es un problema actual de la medicina que se encamina hacia una ciencia menos artística^{1,2}.

Los índices de salud en general se han desarrollado mucho en las últimas décadas. Índices con aplicaciones en la predicción, la discriminación o la evaluación se han propuesto en la literatura reciente³⁻⁶.

Los indicadores que intentan la gravedad de los pacientes durante su estancia en un hospital se han utilizado en particular como variables de control para la evaluación de los servicios hospitalarios^{4,5,7-9}.

En este contexto, una de las escalas de gravedad más estudiadas es la propuesta por Horn y cols. en 1983^{7,10}. Este índice pretendía ser capaz de caracterizar con exactitud las diferencias entre los pacientes por concepto de gravedad y uso de recursos pues debería dar lugar a grupos de pacientes heterogéneos en diagnósticos («case-mix») pero homogéneos en con-

sumo de recursos. La confiabilidad y la validez de este índice fueron abordados en un estudio posterior donde se incluyeron pruebas de confiabilidad realizadas en 18 hospitales (especialmente concordancia entre jueces), validez de contenido y de presentación (sobre la base de opiniones de personal de distintos hospitales) y validez de criterio y construcción¹¹. El índice demostró ser aceptablemente válido y confiable.

Otras formas conocidas de medir o evaluar la gravedad de la enfermedad del paciente hospitalizado son: el sistema de grupos de diagnósticos relacionados (diagnostic related groups, DRG)¹² y el sistema de estadiamiento propuesto por Gonnella y cols¹³. El primero se creó con el fin de clasificar a los pacientes en categorías de costos y servía para implementar un sistema de pagos prospectivos para los pacientes adscritos a Medicare (una agencia norteamericana de seguros médicos) y el segundo pretendía brindar un sistema comprensible de clasificación de la mayoría de diagnósticos según gravedad.

Tanto con los DRG como con el sistema de estadiamiento de Gonnella y otros, la gravedad se le adjudica a un conjunto de características centradas en el diagnóstico y las categorías de gravedad se construyen independientemente de cada paciente. En ambos casos se necesitaron paneles de expertos para construir las combinaciones de los diagnósticos con otras características que definirían los distintos grupos y grado de gravedad.

En Cuba no se ha implantado ninguno de los sistemas conocidos y, por el momento, creemos que no existe ninguna escala que se adapte a nuestras condiciones. La escala propuesta por Horn demanda la existencia de personal especializado y con dedicación completa, o casi completa, a la actividad evaluadora. Los DRG y el sistema de estadiamiento, han sido elaborados a base de información de sus propios países, con condiciones muy diferentes de las nuestras.

Con el presente trabajo pretendemos presentar una manera, razonablemente sencilla y poco costosa de medir la gravedad de la enfermedad, orientada a su introducción en la práctica actual de nuestros hospitales. Se aspira a que el índice o escala resultante cumpla con los requisitos de ser factible y rápido de obtener por un médico a partir de los datos existentes en las historias clínicas y de ser generalizable a todas las áreas de hospitalización de servicios clínicos para adultos (exceptuando los servicios especiales de terapia, quemados y servicios psiquiátricos).

Método

Todo el proceso de construcción y validación del índice se llevó a cabo a partir de entrevistas a especialistas o revisión de historias clínicas de los servicios del

área clínica del hospital «Hermanos Ameijeiras» que incluyen, en orden creciente de tamaño los siguientes: reumatología, endocrinología, neurología, gastroenterología, hematología, nefrología, cardiopatía y medicina interna. Las edades de los pacientes cuyas historias clínicas fueron incluidas en una de las dos fases, oscilaban entre 14 y 80 años.

Construcción

Se consideró que el índice o escala resultante debería estar formado por componentes, aspectos o ítems que se integraran en un índice global que tomaría la forma de una combinación lineal de los ítems:

$$I = w_1X_1 + w_2X_2 + \dots + w_kX_k$$

donde x_i es la i -ésima variable y w_i el peso escogido para dicha variable.

De este modo la fase de construcción estuvo dirigida a: 1) la selección de los ítems que conformarían el indicador y 2) la búsqueda de los pesos o ponderaciones que debería llevar cada ítem en la combinación lineal.

Inicialmente se consideraron distintas variables o ítems incluidas en las escalas conocidas y las sugerencias de especialistas con más de 10 años de experiencia de las áreas clínicas. Se procuró que estas variables tuvieran un carácter dicotómico (presencia o ausencia de determinada característica).

El proceso de selección de los ítems definitivos y búsqueda de las ponderaciones adecuadas para cada uno se llevó a cabo a partir de la información de las historias clínicas de las primeras 74 altas del mes de abril de 1993 del área clínica del hospital. Estas historias aportaron los datos sobre todos los ítems y fueron evaluadas además según el índice de Horn. Este último con dos adaptaciones: se eliminó el aspecto acerca de la dependencia del paciente de la enfermera, porque en las historias clínicas no existía información suficiente para evaluarlo, y se utilizó la suma de los puntos obtenidos en los seis aspectos (evaluados cada uno entre 1 y 4 puntos) en lugar de la moda como fue propuesto originalmente por sus autores.

Se evaluó entonces la asociación entre todos los componentes, para detectar posibles redundancias en la información, y la asociación entre cada componente y el índice de Horn.

En ambos casos se utilizó el coeficiente de correlación de rangos de Spearman. Con cada par de componentes que presentaba una alta correlación (mayor de 0,9), se realizó una valoración casuística para excluir uno de los miembros de la pareja. Se eliminaron las variables que no tenían asociación significativa con el índice de Horn. Consideramos que, con este tamaño

de muestra, una variable que no tuviera correlación significativa con el índice de Horn, no sería un buen exponente de la gravedad.

Las categorías para cada una de las variables que formarían parte de la fase final se muestra en la tabla 1.

Estas variables se incluyeron en un formulario con una explicación detallada sobre el significado de cada categoría. En el anexo se describe la definición operativa de algunas variables.

Para la obtención de las ponderaciones se utilizó la técnica de regresión lineal múltiple. Se escogieron como «pesos» o ponderaciones los coeficientes estandarizados de las variables de la función que se obtenía al considerar el índice de Horn (de 6 a 24 puntos) como variable de respuesta y los ítems como explicativas. Se tomaron estos coeficientes multiplicados por 100 y redondeados a entero.

La aplicación de esta técnica condujo también a la inclusión de la variable «suma de órganos con fallo» en lugar de las cinco dicotómicas que reflejaban la presencia de fallos de órganos por separado (cardíaco, renal, respiratorio, hepático y neurológico) ya que algunos de los fallos estaban presentes en muy pocos pacientes y resultaba importante el hecho de que un paciente tuviera más de un fallo orgánico.

Con ayuda del método de selección de variables paso a paso se arribó a la versión final del indicador que proponemos.

Para calcular el índice en un paciente basta sustituir cada variable por sus valores, de acuerdo a los códigos mostrados en la tabla 1 y evaluar la función con las ponderaciones pertinentes (tabla 2).

Por último se procedió a clasificar la gravedad según una escala ordinal con cuatro niveles de gravedad, recur-

Tabla 1. Escala de las variables introducidas en el modelo inicial

Edad:			
1) 14 a 35 años	2) 36 a 60 años	3) 61 años y más	
Tipo de ingreso:			
0) electiva	1) urgente		
Manifestaciones de la enfermedad:			
0) ninguna	1) no amenaza la vida	2) amenaza la vida	
Localización del diagnóstico principal:			
0) piel, TCS, musculoesquelético o sistema nervioso periférico	1) otra localización (incluye grandes sistemas por separado y sistémico)		
Etiología del diagnóstico principal:			
1) neoplásica o degenerativa	0) otra		
Localización de las enfermedades asociadas:			
1) sistema respiratorio, cardiovascular o SNC	0) otra		
Uso de procedimientos invasivos diagnósticos o terapéuticos:			
0) no	1) sí		
Uso de algún procedimiento urgente:			
0) no	1) sí		
Estuvo en la unidad de cuidados intensivos o intermedios:			
0) no	1) sí		
Complicaciones:			
0) ninguna	1) no amenazaron la vida	2) amenarazon la vida	
Tipo de respuesta al tratamiento:			
0) inmediata	1) no inmediata pero el paciente respondió	2) no respuesta	
Efectos residuales:			
0) ninguno	1) alguno que no cambió hábitos de vida	2) alguno que sí cambió hábitos de vida	3) fallecido
Transfusiones de sangre o derivados:			
0) no	1) sí		
Fallos orgánicos (por separado: cardíaco, renal, respiratorio, hepático, neurológico):			
0) no	1) sí		

Anexo. Significado de algunas componentes del índice de gravedad propuesto incluidos en la tabla 1

Diagnóstico principal:

Condición morbosa, establecida después del estudio del paciente, que realmente motivó el proceso que dio lugar a la hospitalización.

Enfermedad degenerativa:

Enfermedades sistémicas inflamatorias del tejido conjuntivo, de etiología no precisada y pronóstico reservado. Sus exponentes principales son: artritis reumatoide, lupus eritematoso sistémico, esclerodermia y enfermedad mixta del tejido conjuntivo.

Procedimiento invasivo:

Aquellos que utilizan instrumentos o dispositivos que invaden órganos o tejidos y que tienen una probabilidad relativamente alta de complicaciones como la radioterapia, la intubación del tracto gastrointestinal o respiratorio, la toracocentesis, las biopsias, el cateterismo cardíaco, las endoscopias y los tratamientos endovenosos que requieren monitorización estrecha para prevenir complicaciones.

Procedimiento urgente:

Aquel que se realiza con el objeto de evitar la muerte en corto lapso por alguna complicación de la enfermedad o de alguna medida tomada en razón de ésta. Ejemplos: procedimientos de reanimación o resucitación cardiopulmonar, implantación del balón de Blakemore-Sengstaken para suprimir sangramientos de varices esofágicas.

Complicaciones:

Eventos indeseables que surgen como consecuencia directa de la enfermedad principal, de su tratamiento, del proceso diagnóstico o la propia hospitalización.

Enfermedad asociada:

Condición morbosa que no constituye parte del diagnóstico principal ni de sus complicaciones pero que puede tener alguna influencia en la evolución de la enfermedad principal o sus complicaciones.

Efecto residual:

Se refiere al grado en que el paciente muestra aún evidencias de enfermedad o daño en el momento de su alta del hospital.

Fallos orgánicos:

Deben considerarse principalmente los que han sido calificados como tales durante la hospitalización y que se plasman en la historia clínica. En caso de dudas el evaluador debe tener en cuenta los siguientes elementos:

cardíaco: FC < 50/mto. Taquicardia ventricular / fibrilación. Paro cardíaco. IMA.

renal: creatinina sérica > 280 mol/l o 3,5 mg/dl. Diálisis / ultrafiltración.

hepático: bilirrubina > 51 mol/l o 3 mg/dl en ausencia de hemólisis. TGP > 100 U/l.

respiratorio: Frec. resp. < 5/min o > 50/min. Vent. Mec. por 3 o más días o FiO₂ > 0,4 o PEF > 5 mmHg.

neuroológico: Glasgow en ausencia de sedación < 6.

Tabla 2. Ponderaciones para las variables resultantes del proceso de selección

Variabes	Ponderación
Edad	15
Tipo de ingreso	11
Etiología del diagnóstico principal neoplásica o degenerativa	11
Procedimientos diagnósticos o terapéuticos invasivos	12
Suma de fallos orgánicos	28
Complicaciones	15
Efectos residuales	25
Respuesta al tratamiento	18

so adecuado para muchos fines y frecuentemente usado en la literatura.

Para ello se dividió la distribución empírica del índice en cuatro «zonas» enmarcadas por los percentiles 10, 50 y 90. En resumen se arribó a la propuesta de dos alternativas para índice de gravedad: uno que da lugar a una variable con valores entre 15 y 459, prácticamente una variable cuantitativa continua, y otra que

toma valores de 1 a 4 y que puede calificarse como una variable ordinal con 4 categorías de menor a mayor gravedad.

Validación

Se exploraron cuatro aspectos de la validez: presentación, contenido, construcción y criterio^{14,15}. Se calculó el coeficiente de confiabilidad y se examinó la concordancia entre jueces y la consistencia interna.

La confiabilidad, la validez de construcción y la de criterio se evaluaron a partir de un nuevo grupo de 100 historias clínicas consecutivas en la lista de los egresos del mes de mayo de 1993 del mismo hospital.

La validez de presentación y contenido se evaluó conjuntamente mediante una encuesta a 16 expertos (médicos especialistas con más de diez años de experiencia profesional) que se pronunciaron sobre el cumplimiento de las cinco propiedades básicas propuestas por Moriyama¹⁶: 1) razonable y comprensible, 2) sensible a variaciones en el fenómeno que se mide, 3) con

suposiciones básicas justificables e intuitivamente razonables, 4) con componentes claramente definidos y 5) derivable de datos factibles de obtener.

Los expertos se pronunciaron sobre el grado de cumplimiento de cada uno de estos principios según tres posibilidades: nada, moderadamente o mucho.

Para el análisis de la validez de construcción del índice se parte de los siguientes supuestos: 1) la mayor parte de los fallecidos se correspondían con los pacientes a los que el índice calificó como los más graves y 2) el tiempo de estancia de los pacientes que egresaron vivos debía estar correlacionado positivamente con el valor del índice. La asociación entre el índice ordinal y el estado al alta hospitalaria (vivo o fallecido) se evaluó a partir de Chi Cuadrado y el coeficiente de asociación para variables ordinales, Tau C (de interpretación similar al coeficiente de correlación de Pearson¹⁷) y los valores promedio del índice entre egresados vivos y fallecidos mediante la *t* de Student para muestras independientes. Para evaluar la asociación entre el índice cuantitativo y la estancia se emplearon los coeficientes de correlación de Pearson unidos a los diagramas de dispersión correspondientes.

Se evaluó la correlación entre el índice y dos indicadores de gravedad para pacientes hospitalizados para el análisis de la validez de criterio: el índice de gravedad de la enfermedad propuesto por Horn¹⁰ con las adaptaciones descritas anteriormente, y el sistema de estadiamiento propuesto por Gonnella y cols.¹³ del cual se utilizaron solamente las bases generales puesto que no se tuvo acceso al sistema completo.

La asociación con el índice de Horn se evaluó a través del coeficiente de correlación lineal de Pearson.

La asociación entre el índice en cuatro categorías y el sistema de Gonnella y cols. se evaluó mediante la prueba chi cuadrado y el coeficiente Tau B, índice adecuado para cuantificar asociación entre dos variables ordinales con igual número de categorías cada una¹⁷.

Confiabilidad

Se calculó el coeficiente de confiabilidad general y se constató el grado de cumplimiento de dos de sus aspectos: la equivalencia en cuanto a concordancia entre «jueces» y la consistencia interna. Se calculó el coeficiente de confiabilidad (R) según la siguiente expresión.

$$R = \frac{\sigma_{pac}^2}{\sigma_{pac}^2 + \sigma_{obs}^2 + \sigma_{err}^2}$$

donde:

σ_{pac}^2 es la estimación de la varianza entre pacientes.

σ_{obs}^2 es la estimación de la varianza entre jueces u observadores.

σ_{err}^2 es la estimación de la varianza del error (debida al azar).

El coeficiente R brinda una estimación de la parte de la variación total que se debe a los pacientes¹⁴.

La equivalencia se demostró a través de cinco «jueces» (médicos) que aplicaron por separado el cuestionario con las variables seleccionadas a las mismas 100 historias clínicas. Los jueces fueron instruidos brevemente y oralmente por los autores, en particular sobre las respuestas a los ítems: presencia de complicaciones, respuesta al tratamiento y efectos residuales que, a diferencia del resto, están categorizadas en tres niveles. Para la evaluación de la concordancia entre jueces se obtuvo la matriz de correlaciones entre los valores del índice evaluados por todos los posibles pares de jueces.

Con el índice ordinal en cuatro categorías se calculó el coeficiente Kappa ponderado para más de dos jueces y los Kappa para cada uno de los cuatro niveles. La significación estadística se calculó para el Kappa global y para cada uno de los Kappa por categorías¹⁸.

Para la evaluación de la consistencia interna se empleo el índice con sus ítems respectivos y se calcularon fundamentalmente los coeficientes alfa de Cronbach.

Se calculó un alfa de Cronbach general y el alfa que se obtiene al eliminar cada uno de los ítems en relación a su consistencia interna.

La consistencia interna también se evaluó a través de los coeficientes de correlación entre cada ítem y el que se formaría con la suma del resto de los ítems. Se calculó también el coeficiente de determinación (R^2) que surge de poner cada ítem como variable dependiente y el resto de los ítems como independientes. Ambos procedimientos brindan información sobre el grado de correlación entre cada ítem y el resto.

Para los cálculos de los diferentes estadígrafos utilizados se usó el paquete de programas SPSS en su versión 4, excepto para el cálculo de los coeficientes Kappa¹⁸ en que se utilizó un programa confeccionado *ad hoc*.

Resultados

Los resultados de la construcción del índice se resumen en la tabla 2, incluyen los dos procesos de selección de variables y los de búsqueda de la ponderación adecuada para cada una. Las variables que mayor peso aportaron fueron los efectos residuales y la suma de fallos.

La tabla 3 muestra los resultados de someter a la consideración de los expertos las componentes del índice para el análisis de la validez de presentación y contenido. Sólo un experto opinó que no era derivable de datos factibles de obtener. En este aspecto sin embargo, la mayoría se inclinó por la opinión de que el índice cumplía en gran medida con esa condición. También la mayoría de los expertos opinó que el índice era muy razonable y comprensible.

Tabla 3. Resultados de las opiniones de 16 expertos

Criterios	Satisface el criterio					
	Nada		Moderadamente		Mucho	
	n.º	%	n.º	%	n.º	%
Razonable y comprensible	0	0	5	31,2	11	68,8
Sensible a variaciones en el fenómeno que se mide	0	0	10	62,5	6	37,5
Suposiciones básicas justificables	0	0	9	56,2	7	43,8
Componentes claramente definidos	0	0	9	56,2	7	43,8
Derivable de datos factibles de obtener	1	6,2	5	31,2	10	62,5

Para las otras propiedades exploradas, entre el 56 y el 62% de los expertos opinó que se cumplían sólo moderadamente mientras que el resto opinó que el contenido del índice cumplía en gran medida con esas condiciones.

La tabla 4 muestra los resultados de evaluar la asociación entre el índice (ordinal) y el estado en el alta que permite el análisis de la validez por construcción. Se aprecia que la asociación es altamente significativa con un coeficiente de 0,4. El promedio del índice en los vivos es de 56,7 (D E = 29,9) y en los fallecidos de 271,3 (D E = 25,4). La diferencia es significativa e importante ($p < 0,0001$). El índice promedio en los fallecidos es casi 5 veces mayor que en los vivos.

Tabla 4. Relación entre índice de gravedad de la enfermedad y estado al egreso

	Estado al egreso				
	Vivos		Fallecidos		
	N.º	%	N.º	%	
Niveles del índice	1	10	11,4	0	0
	2	43	48,9	0	0
	3	35	39,8	3	25,0
	4	0	0	9	75,0

$\chi^2 = 73,8$ $p < 0,0001$ Tau C = 0,38

La estancia no tiene correlación significativa con el índice obtenido.

En relación a la validez de criterio, la correlación con el índice de Horn es mayor de 0,9. Por otro lado, la asociación entre el índice en cuatro categorías y el sistema de estadiamiento propuesto por Gonnella (tabla 5) es altamente significativa y el coeficiente de asociación mayor de 0,5.

Se observa una tendencia a que las historias evaluadas como 1 o 2 según Gonnella se evalúen más alto según nuestro índice.

El coeficiente R, sus componentes y el análisis de

Tabla 5. Relación entre índice de gravedad de la enfermedad y sistema de estadiamiento de Gonnella

		Sistema de Gonnella (grados)			
		1	2	3	4
Niveles del índice	1	9	1	0	0
	2	31	9	3	0
	3	13	19	3	3
	4	0	0	0	9

$\chi^2 = 89,8$ $p < 0,0001$ Tau B = 0,54

Tabla 6. Análisis de la varianza y coeficiente de confiabilidad para el índice cuantitativo

Fuente de variación	Suma de cuadrados	GL	Cuadrado medio	F	Sign
Efectos principales	1728158,8	103	16778,2	99,9	0,000
Observador	11214,8	4	2803,7	16,7	0,000
Paciente	1716944,0	99	17342,9	103,3	0,000
Explicada	1728158,8	103	16778,2	99,9	0,000
Residual	66510,4	396	168,0		
Total	1794669,2	499	3596,5		

$\sigma^2_{err} = 168$; $\sigma^2_{pac} = 3434,98$; $\sigma^2_{obs} = 26,36$; $R = 0,9464$

varianza original que permitan evaluar la concordancia entre observadores, se muestran en la tabla 6. El valor de R (0,95) indica que la mayor parte de la variación total entre los datos se debe a los pacientes. La variabilidad debida a los observadores es sólo de 0,7%. Los coeficientes de correlación entre todos los pares de observadores están por encima de 0,93. El coeficiente kappa ponderado fue de 0,38. Los kappa por categorías oscilaron entre 0,26 (para la categoría 2) y 0,86

(para la categoría 4). Todos resultaron ser significativamente mayores que 0 ($p < 0,001$).

El coeficiente alfa de Cronbach (tabla 7) para el análisis de la consistencia interna fue de 0,78. Según este procedimiento los ítems más relevantes, ya que de ser eliminados causarían las mayores disminuciones en el coeficiente α , resultaron: el de complicaciones, el de efectos residuales, el de la suma de fallos y el de respuesta al tratamiento. Algunos ítems tienen poca correlación con el resto unidos, tales como la edad, la etiología del diagnóstico principal, el uso de procedimientos invasivos y el tipo de ingreso.

Tabla 7. Elementos de la consistencia interna para el índice después del proceso de selección

VARIABLES	Correlación ítem-total corregida	Coefficiente de determinac.	Alfa al eliminar la variable
• Edad	0,13	0,06	0,80
• Tipo de ingreso	0,35	0,25	0,78
• Etiología del diagnóstico principal neoplásica o degenerativa	0,39	0,22	0,78
• Procedimientos diagnósticos o terapéuticos invasivos	0,33	0,19	0,78
• Suma de fallos orgánicos	0,81	0,71	0,69
• Complicaciones	0,78	0,68	0,72
• Efectos residuales	0,84	0,73	0,69
• Respuesta al tratamiento	0,68	0,54	0,73

Alfa de Cronbach = 0,78; alfa de ítems estandarizados = 0,79

Discusión

El sistema de pesos o ponderaciones presentado aquí ha sido poco utilizado. Streiner y Normal¹⁴ mencionan el uso de la regresión múltiple pero son poco explícitos en cuanto a la forma de aplicación, pues su uso en este contexto implica un círculo vicioso: la variable de respuesta deberá ser una evaluación cuantitativa del aspecto para el que precisamente se necesita la escala que se está construyendo. En el presente trabajo se intenta romper este círculo con un indicador cuantitativo supuestamente válido y confiable pero más complicado de obtener.

La mayoría de los trabajos de este tipo incluyen pesos más o menos arbitrarios obtenidos muchas veces del consenso de grupos de expertos^{13,19,20}. En un estudio reciente, donde se desarrolló un índice de gravedad para evaluar riesgo neonatal y comparar el desempeño de unidades de cuidados intensivos neonatales, se utilizó un método de pesos similar al propuesto, con el modelo de regresión logística⁹.

Si bien el sistema de ponderaciones propuesto no puede considerarse reflejo exacto de la verdadera gravedad, presenta la ventaja de basarse en datos obte-

nidos de manera sencilla, lo que lo hace perfectible y fácilmente adaptable a diferentes entornos. Las ponderaciones encontradas representan razonablemente la cuantificación de la importancia relativa a cada variable. Es de esperar que la presencia de fallo de órganos tenga coeficientes altos, y que variables como el tipo de ingreso y el uso de procedimientos invasivos obtengan coeficientes más bajos.

Los valores exactos de los coeficientes podrían utilizarse sin cambios en el momento actual en el área clínica del hospital «Hermanos Ameijeiras». Su extrapolación a otros centros, sin embargo, demandaría ajustes hechos con información del centro donde se pretenda aplicar. Igualmente, dentro de un mismo hospital, habría que hacer ajustes a las ponderaciones con el trascurso del tiempo pero el procedimiento metodológico escogido podría ser básicamente el mismo.

Los procedimientos de validación utilizados aquí son los que, en términos generales, aconseja la literatura¹⁴, aunque la selección de patrones o puntos de referencia que se utilizan para algunos de los procedimientos validativos es original.

La validez de presentación y la de contenido son los atributos más difíciles de evaluar en índices como éste, pues no existen criterios que puedan cuantificarse claramente. Bohrnstedt²¹ recomienda comenzar este tipo de evaluación por la revisión de la literatura, la construcción del índice que se propone comenzó precisamente por este paso, y por lo tanto, todos sus elementos han estado incluidos en índices construidos y validados con objetivos similares al nuestro.

Teóricamente, la validez de contenido refleja el grado en que se ha muestreado representativamente el dominio que se quiere medir con el indicador. Este enunciado no tiene una consecuencia práctica clara. No obstante, un experto en la materia puede pronunciarse sobre la relación entre lo que para él representa el dominio de estudio y los ítems que aparecen en el indicador. Nuestra intención fue concretar, guiados por las propiedades que propone Moriyama¹⁶, las opiniones de profesionales con experiencia en el manejo de pacientes y de historias clínicas. El resultado puede considerarse muy satisfactorio.

En lo referente a la validez por construcción, el índice presenta una buena relación con el estado del paciente en el momento del alta. Esto está sin duda condicionado por el hecho de que uno de los ítems con mayor ponderación, el de efectos residuales, tiene en su categoría más alta a los fallecidos. Tratándose de un indicador de gravedad del proceso de hospitalización no es desdeñable este hecho. Sin embargo, la estancia no estuvo relacionada con el índice. El resultado es sorprendente pues, tratándose de egresos vivos, se esperaba una asociación positiva ente el índice y la estancia.

Está claro, no obstante, que la estancia no depende solamente de la gravedad, se sabe que buena parte de ella se debe a otras variables como el lugar de resi-

dencia, el diagnóstico principal o la propia eficiencia del trabajo hospitalario. Horn y cols.²², encuentran que la gravedad de la enfermedad, (medida por el índice propuesto por ellos) sólo explicaba el 11% de la longitud de la estancia.

La validez de criterio resultó más sencilla de evaluar. La asociación alta y significativa del índice de Horn con el nuestro podría ser una consecuencia de haber utilizado al primero como base para la obtención de las ponderaciones aunque la construcción se hizo con un grupo de historias clínicas distinto que la validación.

La validez de criterio no es usualmente evaluada en trabajos de esta área y no existe uniformidad alguna entre los indicadores que deben tomarse de referencia ni sobre la magnitud que deben tener las correlaciones para que el índice se considere aceptable.

Feinstein²³ menciona la posibilidad de buscar un patrón estándar para evaluar validez de criterio a base de consenso entre expertos. Podría ser una opción para validaciones futuras del índice propuesto o sus versiones mejoradas.

La confiabilidad, es evaluada con mayor frecuencia y uniformidad. En particular la concordancia entre jueces resulta la característica más deseable para el índice que se propone. El coeficiente de confiabilidad es sumamente alto al igual que los coeficientes de correlación obtenidos de la concordancia entre jueces.

El coeficiente kappa global obtenido con nuestros datos puede considerarse, según criterios de Fleiss¹⁸, aceptable. Para la categoría 4 (la de pacientes más graves) el coeficiente kappa es excelente. Este resultado es relevante porque la mayor parte de las veces lo que más interesa es no cometer errores en la clasificación de los pacientes más graves. Schumacher y cols.²⁴ encuentran coeficientes similares o menores que los nuestros al evaluar este tipo de confiabilidad para el índice de gravedad de Horn y para el índice de sucesos adversos.

Los bajos coeficientes kappa de las categorías 2 y 3 deberán conducirnos en el futuro a probar con alternativas de clasificación de tres categorías de grave-

dad que podrían ser, inclusive, más sencillas y prácticas.

La evaluación de la consistencia interna constituye motivo de controversias. El coeficiente alfa de Cronbach pretende ser un compendio de la correlación entre ítems y se espera que un «buen» índice tenga un alto coeficiente alfa. Sin embargo, si bien pensamos medir un aspecto único (la gravedad del paciente), opinamos que éste resulta de la conjunción de diferentes elementos. No hay que esperar una correlación demasiado intensa entre estos aspectos sino que la suma de ellos brinde un indicador real de la gravedad del paciente en términos globales.

Feinstein²³ señala el poco uso que se le ha dado al alfa de Cronbach en el área de los índices clínicos, quizá por la deliberada construcción de los mismos a base de variables que tienen diferentes roles.

El índice de Horn ha sido criticado por su vulnerabilidad a problemas de calidad de la atención²⁴. El índice que proponemos resulta también vulnerable en este sentido, en particular el uso de procedimientos invasivos podría ser resultado de una indicación errónea y no de una mayor gravedad del paciente. La magnitud de este problema podría evaluarse empíricamente pero, por el momento, no consideramos que esto invalide el uso del indicador.

El proceso de validación mostró un buen comportamiento del índice que le permite comenzar a ser aplicado en el área de servicios clínicos. No obstante, consideramos que sería conveniente mejorar algunos aspectos en el futuro.

Concretamente, sugerimos: evaluar la posibilidad de otras alternativas de clasificación ordinal y el efecto de considerar entre sus componentes alguno que también depende de la calidad de la atención. Recomendamos continuar el proceso de validación con grupos nuevos de historias clínicas y de evaluadores. Una de las ventajas principales de este sistema de clasificación de pacientes es la facilidad con que admite modificaciones una vez implementado su uso rutinario.

Bibliografía

1. Pozo F. La práctica clínica y la epidemiología clínica. Cuadernos de Perseo 1991;1(1):20-1.
2. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: W.B. Saunders Company; 1985.
3. Kirshner B, Guyat G. A methodological framework for assessing health indices. J Chron Dis 1985;38(1):27-36.
4. The international neonatal network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. Lancet 1993;342:193-8.
5. Copeland GP, Jones D, Wilcox A y cols. Comparative vascular audit using the POSSUM scoring system. Am R Coll Surg Eng 1993;75:175-7.
6. Rabeneck L, Feinstein AR, Horwitz RI, Wells CK. A new clinical prognostic staging system for acute pancreatitis. Am J Med 1993;95(1):61-70.
7. Horn SD. Measuring severity of illness: comparisons across institutions. Am J Publ Health 1983;73:25-31.
8. Rosko MD, Carpenter CE. Development of a scalar hospital-specific severity of illness measure. J Med Syst 1993;17(1):25-36.
9. Taroni F, Louis DZ, Yen EJ. An analysis of health services using disease staging: a plot study in the Emilia-Romagna region of Italy. Journal of Management in Medicine 1992;6(2):53-66.
10. Horn SD, Sharkey PD, Bertram DA. Measuring severity of illness: homogeneous case mix groups. Med Care 1983;21(1):14-30.
11. Horn SD, Horn RA. Reliability and validity of the severity of illness index. Med Care 1986;24(2):159-78.

12. Fetter RB, Shin Y, Freeman JL, Averill R, Thompson J. Case mix definitions by diagnosis-related groups. *Med Care* 1980;19(Supl 1):1-53.
 13. Gonnella JS, Hornbrook MC, Louis DZ. Staging of disease. *JAMA* 1984;251(5):637-44.
 14. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press; 1989.
 15. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: McGraw-Hill Book Company; 1985.
 16. Moriyama IM. Indicators of social change. En: Shelden E, Moore W, eds. Problems in the measurement of health status. New York: Russel Sage Foundation; 1968.
 17. Nie NH, Hull CH, Jenkins JG y cols. Statistical package for the social sciences. New York: McGraw-Hill Book Company; 1975.
 18. Fleiss JL. Statistical methods for rates and proportions. 2 ed. New York: John Wiley and Sons; 1981.
 19. Gustafson DH, Fryback DG, Rose JH y cols. An evaluation of multiple trauma severity indices created by different index development strategies. *Med Care* 1983;21(7):674-91.
 20. Knaus WA, Zimmerman JE, Wagner DP y cols. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;9:591-7.
 21. Bohrnstedt GW. Measurement. En: Handbook of survey research. London: Academic Press; 1983.
 22. Horn SD, Sharkey PD, Buckle JM y cols. The relationship between severity of illness and hospital length of stay and mortality. *Med Care* 1991;29:305-17.
 23. Feinstein AR. Clinimetrics. New Haven: Yale University Press; 1987.
 24. Schumacher DN, Parker B, Kofie V y cols. Severity of illness index and the adverse patients occurrence index. A reliability study and policy implications. *Med Care* 1987;25:695-704.
-