

Estudio primario

Medwave 2017 Mar;17(2):e6880 doi: 10.5867/medwave.2017.02.6880

Índice de gravedad para servicios quirúrgicos en un hospital cubano: estudio de revalidación

Severity of illness index for surgical departments in a Cuban hospital: revalidation study

Autores: Amadys Armas-Bencomo[1], Teddy Osmin Tamargo-Barbeito[2], Edelberto Fuentes-Valdés[3], Rosa Eugenia Jiménez-Paneque[2]

Filiación:

[1] Departamento de Admisión, Archivo y Estadística, Centro de Investigaciones Médico Quirúrgico (CIMEQ), La Habana, Cuba

[2] Departamento de Investigaciones y Proyectos, Hospital Hermanos Ameijeiras, La Habana, Cuba

[3] Servicio de Cirugía General, Hospital Hermanos Ameijeiras, La Habana, Cuba

E-mail: amadimb@gmail.com

Citación: Armas-Bencomo A, Tamargo-Barbeito TO, Fuentes-Valdés E, Jiménez-Paneque RE. Severity of illness index for surgical departments in a Cuban hospital: revalidation study. *Medwave* 2017 Mar;17(2):e6880 doi: 10.5867/medwave.2016.02.6880

Fecha de envío: 7/7/2016

Fecha de aceptación: 26/1/2016

Fecha de publicación: 8/3/2017

Origen: no solicitado

Tipo de revisión: con revisión por tres pares revisores externos, a doble ciego

Palabras clave: severity of illness index, validity, reliability, risk adjustment

Resumen

INTRODUCCIÓN

En el contexto de la evaluación de servicios hospitalarios, la incorporación de índices de gravedad permite tener una variable de control esencial para la comparación del desempeño en el tiempo y el espacio a través del llamado ajuste por riesgo. El índice de gravedad para servicios quirúrgicos, se construyó en 1999 y se validó como un índice general para estos servicios. Dieciséis años después el contexto hospitalario ha cambiado y se consideró necesaria una revalidación de este índice de gravedad que avale su utilidad actual.

OBJETIVO

Evaluar la validez y confiabilidad del índice de gravedad para servicios quirúrgicos, que avale su uso razonable en las condiciones actuales.

MÉTODOS

Se realizó una investigación descriptiva retrospectiva en el servicio de cirugía general del Hospital Clínico Quirúrgico Hermanos Ameijeiras en el segundo semestre del año 2010. Se revisaron las historias clínicas de 511 pacientes egresados de este servicio. Las variables utilizadas fueron las mismas del índice original con sus ponderaciones. Se evaluaron validez conceptual o de constructo, validez de criterio y confiabilidad interevaluadores así como consistencia interna del índice propuesto.

RESULTADOS

La validez de constructo se expresó en una asociación significativa entre el valor del índice de gravedad para servicios quirúrgicos y el estado al egreso. Se comprobó también correlación significativa, aunque débil, con la estadía hospitalaria. En cuanto a validez de criterio, la correlación entre el índice de gravedad propuesto y el índice de Horn fue de 0,722 (intervalo de confianza de 95%: 0,677-0,761); mientras que con el índice Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (POSSUM) la correlación fue de 0,454 (intervalo de confianza de 95%: 0,388-0,514) con

el riesgo de muerte y 0,539 (intervalo de confianza de 95%: 0,462-0,607) con el riesgo de morbilidad. La consistencia interna mostró α de Cronbach estandarizado de 0,8; la confiabilidad interevaluadores resultó en un coeficiente de confiabilidad de 0,98 para el índice de gravedad para servicios quirúrgicos cuantitativo y un coeficiente de κ ponderado global de 0,87 para el índice de gravedad para servicios quirúrgicos ordinal.

CONCLUSIONES

La validez y confiabilidad del índice propuesto fue adecuada en todos los aspectos evaluados. El índice de gravedad para servicios quirúrgicos puede utilizarse en el contexto original y es fácilmente adaptable a otros contextos.

Abstract

INTRODUCTION

In the context of the evaluation of hospital services, the incorporation of severity indices allows an essential control variable for performance comparisons in time and space through risk adjustment. The severity index for surgical services was developed in 1999 and validated as a general index for surgical services. Sixteen years later the hospital context is different in many ways and a revalidation was considered necessary to guarantee its current usefulness.

OBJECTIVE

To evaluate the validity and reliability of the surgical services severity index to warrant its reasonable use under current conditions.

METHODS

A descriptive study was carried out in the General Surgery service of the "Hermanos Ameijeiras" Clinical Surgical Hospital of Havana, Cuba during the second half of 2010. We reviewed the medical records of 511 patients discharged from this service. Items were the same as the original index as were their weighted values. Conceptual or construct validity, criterion validity and inter-rater reliability as well as internal consistency of the proposed index were evaluated.

RESULTS

Construct validity was expressed as a significant association between the value of the severity index for surgical services and discharge status. A significant association was also found, although weak, with length of hospital stay. Criterion validity was demonstrated through the correlations between the severity index for surgical services and other similar indices. Regarding criterion validity, the Horn index showed a correlation of 0.722 (95% CI: 0.677-0.761) with our index. With the POSSUM score, correlation was 0.454 (95% CI: 0.388-0.514) with mortality risk and 0.539 (95% CI: 0.462-0.607) with morbidity risk. Internal consistency yielded a standardized Cronbach's alpha of 0.8; inter-rater reliability resulted in a reliability coefficient of 0.98 for the quantitative index and a weighted global Kappa coefficient of 0.87 for the ordinal surgical index of severity for surgical services (IGQ).

CONCLUSIONS

The validity and reliability of the proposed index was satisfactory in all aspects evaluated. The surgical services severity index may be used in the original context and is easily adaptable to other contexts as well.

Introducción

Los indicadores de resultados (*outcomes*) de la atención sanitaria (mortalidad, complicaciones, reingreso y otros) constituyen, junto con los indicadores de estructura y de proceso, los pilares de la evaluación del desempeño de los servicios de salud [1],[2]. La mayoría de las veces esta evaluación conlleva algún tipo de comparaciones con otros servicios o con otros momentos del tiempo. Sin embargo, las comparaciones con indicadores de resultados deben tener en cuenta las características de los pacientes que incluyen, en particular aquellas que podrían estar relacionadas con el resultado en evaluación. Esto significa

controlar variables potencialmente influyentes en los resultados y que no son objeto de evaluación directa, un procedimiento que se conoce como "ajuste por riesgo" (en inglés *risk adjustment*) [3],[4].

La necesidad del ajuste por riesgo se comprobó por primera vez en 1863 cuando Florence Nightingale calificó de peligrosos a los hospitales de Londres, a partir de los datos de mortalidad publicados en ese entonces. En ellos aparecían abismales diferencias entre la mortalidad de hospitales londinenses y la de hospitales rurales [5]. La

percepción de F. Nightingale resultó una falacia, porque justamente la comparación no estaba “ajustada por riesgo”. Hoy se trata de un tema ampliamente explorado, como queda recogido en el libro editado por Lisa Iezzoni por primera vez en 1994 (*Risk Adjustment for Measuring Healthcare Outcomes*) que llega ya a su cuarta edición [6], y en numerosos artículos posteriores [7],[8].

La necesidad de ajuste por riesgo se extiende a cualquier comparación entre servicios de salud de cualquier nivel (salas, hospitales, servicios de especialidades), con la intención de evaluar desempeño a través de indicadores de resultados como mortalidad, morbilidad o tiempo de estadía. Se han desarrollado numerosos sistemas de ajuste por riesgo que se pueden clasificar según diferentes ejes [4]. Una forma de clasificarlos distingue los que se construyen sobre la base de paneles de expertos que dan lugar a agrupaciones diagnósticas -que permiten formar estratos y hacer comparaciones estratificadas- de los que se fundan en características de los pacientes, que pueden incorporarse a funciones de regresión con variables de resultados como variables dependientes. Entre los primeros se destacan el sistema de los *Diagnostic Related Groups* y el *Disease Staging* que, por la forma de construirlos, se han convertido en sistemas comerciales y se utilizan particularmente en los sistemas de pagos de la asistencia sanitaria [9],[10].

Otra forma de clasificación distingue los que se aplican a enfermedades o procedimientos específicos y los que se consideran genéricos, porque se aplican a servicios de salud enteros. En la evaluación de desempeño de los servicios de salud, hospitales en particular, se necesitan formas de ajuste que permitan comparaciones globales en tiempo y espacio independiente de los grupos conformados por sistemas comerciales como los mencionados antes. Las más socorridas son combinaciones de características de los pacientes y su enfermedad que se asocian al riesgo de determinados resultados y suelen englobarse en el concepto de “gravedad de la enfermedad” (*severity of illness*) [11],[12].

Se han propuesto múltiples índices de gravedad de la enfermedad específicos para distintos servicios o especialidades, o incluso para determinadas enfermedades. Pueden ser útiles en la asistencia médica (como indicadores de pronóstico), pero también se utilizan para el ajuste por riesgo. En las unidades de cuidados intensivos se destaca, por ejemplo, el sistema *The Acute Physiology and Chronic Health Evaluation* (APACHE), y otros semejantes como *Mortality Prediction Model* (MPM) y el *The Simplified Acute Physiology Score* (SAPS) [13].

Se han desarrollado en distintas latitudes índices de gravedad con el objetivo de realizar, de alguna manera, un ajuste por riesgo. Sin embargo, son más bien pocos los índices de gravedad de aplicación general o para una amplia gama de servicios en un hospital. Una de las escalas más generales es la que propusieron Horn y colaboradores en 1983 y que llamaron *Severity of Illness Index* [14],[15]. Este índice contiene siete dimensiones de la asistencia médica, cada una de las cuales se evalúa en cuatro niveles

desde el que implica menor gravedad (nivel 1) hasta el que se corresponde con mayor gravedad (nivel 4). Es aplicable a pacientes hospitalizados de cualquier servicio clínico, quirúrgico o cuidados intensivos, permite formar agrupaciones de pacientes que pueden ser heterogéneos en cuanto a diagnóstico, pero homogéneos en el consumo de los recursos hospitalarios. Sin embargo, por lo que parece, fue poco utilizado. Su limitación principal es la necesidad de entrenamiento de los evaluadores para lograr confiabilidad aceptable.

Pero el ajuste por riesgo no es solo necesario para el trabajo hospitalario administrativo, sino una necesidad de cualquier comparación o evaluación que se haga a partir de algún indicador de resultados. Hall y colaboradores en la revista *Surgery* hacen una revisión extensa del uso del ajuste por riesgo en toda la investigación observacional, particularmente la que se hace con datos de registros [16]. Incluso en los ensayos clínicos, donde la aleatorización supuestamente evita toda diferencia entre los grupos, el ajuste por riesgo puede ser importante si se quieren tamaños de efecto más informativos y que contemplen las diferencias entre pacientes dentro de los grupos.

Como ya se esbozó antes, las escalas o índices de gravedad de la enfermedad (*severity of illness indexes*) resultan los mediadores más socorridos para el menester de ajustar por riesgo las comparaciones, al unir en una sola variable la información de todas las que conforman el nivel de gravedad o “riesgo” de tener cierto resultado. Además del ya mencionado y quizás el primero *Severity of Illness Index* propuesto por Horn y colaboradores en 1983, otros han surgido después de él [17],[18].

Un índice de este tipo en el área de la cirugía, que también puede considerarse genérico, es el *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* (POSSUM), desarrollado en el año 1991 y que está orientado específicamente a pacientes quirúrgicos. Se ha utilizado ampliamente en los últimos años para predecir la mortalidad y la morbilidad en diferentes y variados procesos quirúrgicos, además de ser una herramienta útil para el ajuste por riesgo en la comparación entre hospitales, servicios o incluso cirujanos [19].

En general, la construcción de estos índices se basa en la metodología propuesta y conocida para su construcción y se acompaña, como es necesario, del proceso de validación correspondiente. La construcción conlleva al menos un proceso de identificación de ítems, uno de ponderación de los mismos y otro de establecimiento del puntaje final. El proceso posterior de validación es ineludible, puesto que todos los pasos anteriores implican cierto grado de subjetividad y no existen criterios de verdad indiscutibles. Sin embargo, es natural que con el transcurso del tiempo las circunstancias y condiciones en las que se construyó, e incluso validó, cierto índice o escala de este tipo, sufran cambios que interfieran con su validez. Por tanto, las revalidaciones son la mejor manera de demostrar la estabilidad del índice creado y su posible permanencia como indicador válido de la gravedad del paciente y su hospitalización.

Con el propósito de utilizar indicadores de resultados ajustados por gravedad en las comparaciones espaciales (entre servicios) y temporales, en el Hospital Hermanos Ameijeiras de La Habana, se construyeron y validaron en los años noventa dos índices de gravedad: uno para servicios clínicos (excluyendo psiquiatría) y uno para servicios quirúrgicos [20],[21]. La utilidad de estos índices fue evaluada después para el ajuste de la estadía hospitalaria y la propuesta de un indicador de ineficiencias en la atención hospitalaria [22]. Los índices mencionados, después de más de una década de su creación, no habían sido sujeto de nuevas validaciones hasta el año 2013. Ese año se reevaluó la validez y la confiabilidad del Índice de Gravedad para Servicios Clínicos en el servicios de reumatología del Hospital Clínico Quirúrgico Hermanos Ameijeiras [23].

El índice para servicios quirúrgicos se trata de una escala cuyo valor se obtiene con datos relativos al paciente, a la intervención quirúrgica a la que fue sometido y a las complicaciones que hayan surgido durante la hospitalización. Como ya se ha mencionado se trata de un índice para realizar ajuste por riesgo en comparaciones a posteriori, es decir, después de concluida la hospitalización. El índice debe dar cuenta de la gravedad del paciente durante su hospitalización ya pasada.

Este índice de gravedad se construyó en 1999 y se validó como un índice general para servicios quirúrgicos. Dieciséis años después, el contexto de los servicios quirúrgicos en el hospital ha variado de diversas maneras. Ha habido cambios en los equipos de trabajo, organización, infraestructura y tecnología disponible. Por tanto, se consideró apropiada su revalidación.

El objetivo del estudio cuyos resultados se presentan aquí, fue reevaluar la validez y confiabilidad del índice de gravedad para servicios quirúrgicos en el servicio de cirugía general de este hospital, que permita su uso razonable en las condiciones actuales del hospital donde se elaboró y otorgue un aval para su generalización a otros centros hospitalarios de Cuba y otros países.

Métodos

Se realizó un estudio descriptivo en el servicio de cirugía general del Hospital Hermanos Ameijeiras durante el período comprendido entre los meses de julio a diciembre de 2010.

La muestra la conformaron las historias clínicas de todos los pacientes egresados del servicio de cirugía general del Hospital Hermanos Ameijeiras en La Habana, durante el período comprendido entre los meses de julio a diciembre de 2010. En el caso de pacientes con varios egresos en este período, se tomaron solo los datos del último. Se excluyeron a pacientes extranjeros y no se incluyeron historias clínicas de pacientes que, por alguna razón, no fueron intervenidos quirúrgicamente. La fuente de información de donde se recogieron los datos fue la historia clínica de cada paciente. Ésta se tomó del archivo de historias clínicas del hospital. La extracción de los datos de las historias la realizaron dos de los investigadores en conjunto, las dudas se consultaban con el autor especialista en cirugía general.

Descripción del índice de gravedad para servicios quirúrgicos

El índice evaluado consta de 15 componentes o ítems. Cada componente se evalúa en la historia clínica según ciertos criterios y después se pondera para obtener el puntaje final. La Tabla 1 muestra la configuración del índice.

Ítem (variable)	Forma de medición y puntos iniciales	Valor de ponderación
Edad	Años cumplidos	0,3
Etiología del diagnóstico principal	No neoplásica: 1; neoplásica: 2.	0,4
Etiología de las enfermedades asociadas	Ninguna: 0; no neoplásica: 1; Neoplásica: 2.	0,7
Localización de enfermedades asociadas	Ninguna: 0; respiratorio, cardiovascular o sistema nervioso central: 2; otra: 1.	0,1
Intervención	Según grupo: 1,2,3, 4 o 5	0,2
Tipo de intervención	Electiva : 0; urgente: 1	0,9
Complicaciones de la herida quirúrgica	No: 0; sí: 1	0,4
Complicaciones generales	No: 0; sí: 1	1,3
Re intervención	No: 0; sí: 1	0,3
Procederes invasivos	No: 0; sí: 1	0,8
Procederes urgentes	No: 0; sí: 1	2,3
Estancia en UCI*	No: 0; sí: 1	0,8
Fallo de órgano	No: 0; sí: 1	1,2
Efectos residuales	Ninguno: 0; alguno que no cambió hábitos de vida: 1; alguno que sí cambió hábitos de vida: 2; fallecido : 3.	3,7
Transfusiones	No: 0; sí: 1	0,1

*UCI: unidad de cuidados intensivos.

Tabla 1. Configuración del Índice de gravedad para servicios quirúrgicos.

Significado de las variables consideradas en el índice.

Se considera diagnóstico principal aquella condición morbosas que es responsable directa de la admisión del paciente en el hospital.

Se toman como enfermedades asociadas otras condiciones morbosas independientes del diagnóstico principal y sus complicaciones.

La intervención se consideró en cinco grupos según su complejidad desde las más complejas (grupo I) hasta las menos complejas (grupo V). Estos grupos fueron propuestos por el especialista cirujano de 30 años de experiencia coautor de este trabajo. Ejemplos de las intervenciones incluidas en cada grupo se muestran en el artículo original [21].

Se consideran complicaciones de la intervención todos los eventos desfavorables que puedan ocurrir y estén relacionados de modo más o menos directo con la intervención quirúrgica realizada. Como complicaciones relacionadas con la herida se incluyen hematomas, dehiscencias, sangramientos, secreción serosa, sepsis de la herida, enfisema subcutáneo, fístulas, edemas, entre otras. Como complicaciones generales se incluyen retención urinaria, cuadro febril, flebitis, sepsis urinaria, náuseas y/o vómitos, distensión abdominal, diarreas, oclusión intestinal, disminución de la hemoglobina, íctero, íleo paralítico, neumonía, empiema, desbalance hidroelectrolítico o ácido básico, sepsis generalizada, trastornos del ritmo cardíaco, shock séptico,

tromboembolismo pulmonar, infarto del miocardio agudo, sangramiento digestivo, etcétera.

Entre los procederes invasivos diagnósticos y terapéuticos se incluyen aquellos procederes que pueden ser considerados cruentos para el paciente, tales como radioterapia, endoscopia, cateterismo cardíaco, terapia respiratoria, intubación gastrointestinal, terapia endovenosa que requiere monitoreo constante para prevenir complicaciones, entre otras. Se excluye la intervención y los procederes relacionados directamente con el acto quirúrgico.

Como procedimientos de urgencia se incluyen aquellos que son indicados sobre una base de urgencia, con el fin de mantener al paciente con vida (excluyéndose la intervención); por ejemplo resucitación cardíaca, colocación de sonda de Sengstaken Blakemore para contener sangramiento digestivo, etcétera.

Se considera que hubo algún fallo orgánico cuando la historia clínica reflejaba diagnóstico de fallo de algún órgano realizado por especialistas de la unidad de cuidados intensivos. También se le consideró fallo orgánico cuando se podía identificar la presencia de falla en al menos uno de los órganos principales de la economía, según los criterios seguidos por la unidad de cuidados intensivos del hospital. Estos criterios son los siguientes:

- **Hepático:** bilirrubina más de 51 moles por litro o tres miligramos por decilitro en ausencia de hemólisis. TGP mayor de 100 unidades por litro
- **Cardíaco:** frecuencia cardíaca menor de 50 latidos por minuto; taquicardia ventricular/fibrilación; paro cardíaco; infarto agudo de miocardio.
- **Renal:** creatinina sérica mayor a 280 moles por litro o 3,5 miligramos por decilitro. Necesidad de diálisis/ultrafiltración.
- **Respiratorio:** frecuencia respiratoria menor de 5 o mayor de 50 respiraciones por minuto. Ventilación mecánica por tres o más días, porcentaje fraccional de oxígeno inspirado menor a 0,4 o presión positiva al final de la espiración menor a cinco milímetros de mercurio.
- **Neurológico:** Escala de Glasgow en ausencia de sedación menor de seis.
- **Digestivo:** presencia de úlceras de estrés que requieran transfusión de más de dos unidades de sangre en 24 horas; colecistitis no litiasica, enterocolitis necrotizante; perforación de intestino.
- **Hematológico:** hematocrito menor de 20%, conteo de leucocitos menor de $0,3 \times 10^9$ por litro, conteo de plaquetas menor de 50×10^9 por litro.

Los efectos residuales se refieren a secuelas que se espera se mantendrán después del período normal de cicatrización de la herida quirúrgica en término de síntomas, signos, evidencias radiológicas o de laboratorio, así como limitaciones funcionales. Se incluyeron en la categoría ninguno (valor 0), aquellos casos en que no queda secuela alguna de la enfermedad después de la operación, por ejemplo timentomía simple. La segunda categoría (valor 1) incluyó casos en los que la operación dejó alguna secuela, pero esta no implicó cambios en los hábitos de vida del paciente por ejemplo una cuadrantectomía de mama. La tercera categoría (valor 2) se refiere a secuelas de mayor envergadura que ocasionan cambios en los hábitos de vida, por ejemplo una gastrectomía subtotal. Los pacientes fallecidos tuvieron valor 3 en esta variable.

El índice considera, además, si en algún momento de su estadía el paciente tuvo que ser trasladado a la Unidad de Cuidados Intensivos o si fue necesario transfundirle sangre.

El puntaje final toma valores entre 5,1 para una persona con 15 años y el resto de las variables en el mínimo valor posible y 57,6 para una persona de 90 años que fallece a consecuencia de una operación con todas las variables en su valor mayor o más complejo.

Se tomaron de la historia los datos para calcular el índice de gravedad para servicios quirúrgicos y este se evaluó en dos versiones. Primero en su forma cuantitativa (índice de gravedad para servicios quirúrgicos cuantitativo) y después en tres categorías, de menor a mayor gravedad (índice de gravedad para servicios quirúrgicos ordinal). Para obtener este último, se calcularon los percentiles empíricos de esta variable y cada paciente se clasificó en una de las tres categorías dadas por los percentiles:

1. Menor que el 25 percentil.
2. Entre el 25 y el 75 percentil
3. Mayor que el 75 percentil.

A partir de los datos de la historia clínica, y para evaluar la característica psicométrica del índice propuesto (como se describe más adelante), se tomaron también datos para calcular en cada paciente el nivel de gravedad según dos escalas que se describen en la literatura y que se introducen a continuación.

1.1. Adaptación del Índice de gravedad de la enfermedad (*Severity of Illness Index*) propuesto por Horn

La información básica se recogió según la tabla que proponen sus autores [24]. A la forma en que se realiza la puntuación original, se le hicieron dos adaptaciones. Primero, se eliminó la variable dependencia del paciente del personal del hospital y sus facilidades ya que en las historias clínicas no se refleja información suficiente para evaluar este punto. Luego se consideró como puntaje global del índice, la suma de los puntos obtenidos en las seis variables después de evaluar cada aspecto con una escala entre 1 y 4 puntos según sus autores. Por tanto, el valor del Índice de gravedad de la enfermedad osciló entre 6 y 24 puntos (de menor a mayor gravedad).

1.2. Evaluación de la escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity*

En esta escala se consideran dos tipos de puntuaciones: una fisiológica, con 12 factores, y la otra de gravedad operatoria, que utiliza seis factores. Con éstas se estima, a través de una ecuación de regresión logística, el riesgo de mortalidad y de morbilidad. Ambos son los indicadores de gravedad que emplea este sistema. Los detalles se describen en el artículo original y en otros posteriores [25].

Evaluación de las propiedades psicométricas del índice de gravedad para servicios quirúrgicos, análisis estadístico

Validez conceptual (de constructo)

Se evaluó en primer lugar midiendo el grado de asociación entre el índice de gravedad para servicios quirúrgicos y dos indicadores hipotéticamente relacionados con la gravedad: el estado al egreso (vivo o fallecido) y la estadía hospitalaria. La asociación entre el índice de gravedad para servicios quirúrgicos cuantitativo y el estado al egreso se realizó comparando los promedios de dicho índice entre vivos y fallecidos, mediante la prueba U de Mann-Whitney para muestras independientes pues solo se tenían 27 fallecidos, algo que no permitía aplicar un método paramétrico para esa comparación.

La asociación entre el índice en su forma ordinal y el estado al egreso se evaluó con el estadígrafo X^2 y el coeficiente de asociación para variables ordinales τC . La relación entre el índice de gravedad para servicios quirúrgicos y la estadía hospitalaria (en los pacientes egresados vivos), se evaluó mediante el coeficiente de correlación de Pearson.

Validez de criterio:

Se realizó mediante la evaluación de la relación entre el índice de gravedad para servicios quirúrgicos cuantitativo y otros dos indicadores de gravedad para pacientes hospitalizados. Se trata del índice de gravedad de *Horn* y el *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* por medio del coeficiente de correlación de Pearson; ambos mencionados antes.

Consistencia interna

Para la evaluación de la consistencia interna se empleó el índice con sus ítems respectivos y se calcularon fundamentalmente los coeficientes α de Cronbach. Se calculó un α de Cronbach general (con y sin estandarización de ítems) y el α que se obtiene al eliminar cada uno de los ítems en relación con su consistencia interna. También se calcularon los coeficientes de correlación entre cada ítem y el que se formaría con la suma del resto de los ítems ponderados, junto con el coeficiente de determinación (R^2) que surge de utilizar cada ítem como variable dependiente, y el resto de los ítems como independientes. Ambos procedimientos brindaron información sobre el grado de correlación entre cada ítem y el resto.

Concordancia interobservadores

Se eligieron 50 historias clínicas de pacientes al azar entrelas 511 que entraron en el estudio. Este es un número de observaciones suficiente para intervalos de confianza de 90 o 95% con precisión absoluta entre 0,2 y 0,5 suponiendo valores previos del coeficiente de correlación intraclase de 0,9 y de 0,8 para κ [26],[27]. Tres evaluadores independientes (autores) obtuvieron el índice de gravedad

para servicios quirúrgicos en cada historia. Con estos datos se calculó el coeficiente de correlación intraclase o coeficiente de confiabilidad con el índice de gravedad para servicios quirúrgicos cuantitativo y la concordancia interobservadores, con el índice de gravedad para servicios quirúrgicos ordinal. Esta última se evaluó con el coeficiente κ ponderado para más de dos jueces y más de dos categorías, y los κ para cada uno de los niveles. La significación estadística se calculó para el κ global y para cada uno de los κ por categorías. También se obtuvieron intervalos de confianza para estos coeficientes.

Aspectos éticos

El estudio fue aprobado por el Consejo Científico del Hospital Hermanos Ameijeiras, donde se realizó el estudio, previa presentación del proyecto en una reunión ordinaria de este consejo. La confidencialidad de la información se garantizó eliminando todo antecedente que pudiera identificar a los pacientes, una vez obtenidos los datos de las historias, y después de confeccionar y validar la base de datos original.

Resultados

De un total de 574 egresos del servicio en ese período hubo 37 pacientes con ingresos repetidos, de los cuales solo se tomó el último; 15 no se encontraban en el archivo; tres pertenecían a pacientes extranjeros y ocho de pacientes a los que no se les realizó ningún procedimiento quirúrgico. Ninguna historia clínica tenía datos faltantes. Se incluyeron entonces 511 pacientes egresados. En la Tabla 2 se muestra la descripción de las principales características de los pacientes incluidos.

Variable	Categorías	Frecuencia	Porcentaje
Edad agrupada	De 14 a 35	55	10,8
	De 36 a 60	274	53,6
	61 y más	182	35,6
Sexo	Femenino	329	64,4
	Masculino	182	35,6
Etiología del diagnóstico principal	No neoplásica	298	58,3
	Neoplásica	213	41,7
Etiología de enfermedades asociadas	Ninguna enfermedad asociada	147	28,8
	No neoplásica	331	64,8
	Neoplásica	33	6,5
Localización de enfermedades asociadas	Ninguna	147	28,8
	Respiratorio, cardiovascular o SNC	268	52,4
	Otra	96	18,8
Grupo de intervención*	I	47	9,2
	II	63	12,3
	III	99	19,4
	IV	180	35,2
	V	122	23,9
Tipo de intervención	Electiva	402	78,7
	Urgente	109	21,3
Complicaciones de la intervención		46	9
Complicaciones generales		64	12,5
Necesidad de re intervención		28	5,5
Empleo de procedimientos invasivos		328	64,2
Empleo de procedimientos urgentes		27	5,3
Estancia en UCI o UCIM		56	11
Algún fallo orgánico		36	7
Efectos residuales	Ninguno	132	25,8
	Alguno que no cambió hábitos de vida	231	45,2
	Alguno que sí cambió hábitos de vida	121	23,7
	Fallecido	27	5,3
Necesidad de sangre	Sí	62	12,1
Total		511	100

*Ver detalles en la referencia [21].
 SNC: sistema nervioso central.
 UCI: unidad de cuidados intensivos.
 UCIM: Unidad de cuidados intermedios.

Tabla 2. Características de los pacientes incluidos en la validación.

La distribución de los valores del índice de gravedad para servicios quirúrgicos cuantitativo no dista de la distribución Normal. La prueba de Kolmogorov Smirnov no rechazó la hipótesis de que la distribución de esta variable es Normal con media 23,08 y desviación estándar 7,036 ($p=0,684$). Los percentiles 25, 50 y 75 empíricos encontrados en esta distribución son los siguientes:

Percentil 25 = 17,9
 Percentil 50 (Mediana) = 23,0
 Percentil 75 = 27,3

Estos percentiles constituyeron los límites para el índice ordinal de tres categorías.

Validez conceptual (o constructo)

La media (\pm desviación estándar) del índice de gravedad para servicios quirúrgicos para los pacientes egresados vivos fue de $22,3 \pm 6,3$ mientras que para los fallecidos fue de $6,5 \pm 6,0$. La diferencia entre ambas medias fue altamente significativa ($p < 0,0001$) con un intervalo de confianza de 95% para la diferencia entre 11,7 y 16,6.

La Figura 1 muestra la relación entre el índice de gravedad para servicios quirúrgicos cuantitativo y la estadía hospitalaria en los pacientes egresados vivos. El coeficiente de correlación de Pearson (r) fue relativamente bajo (0,32) pero significativamente distinto de 0 ($p < 0,001$).

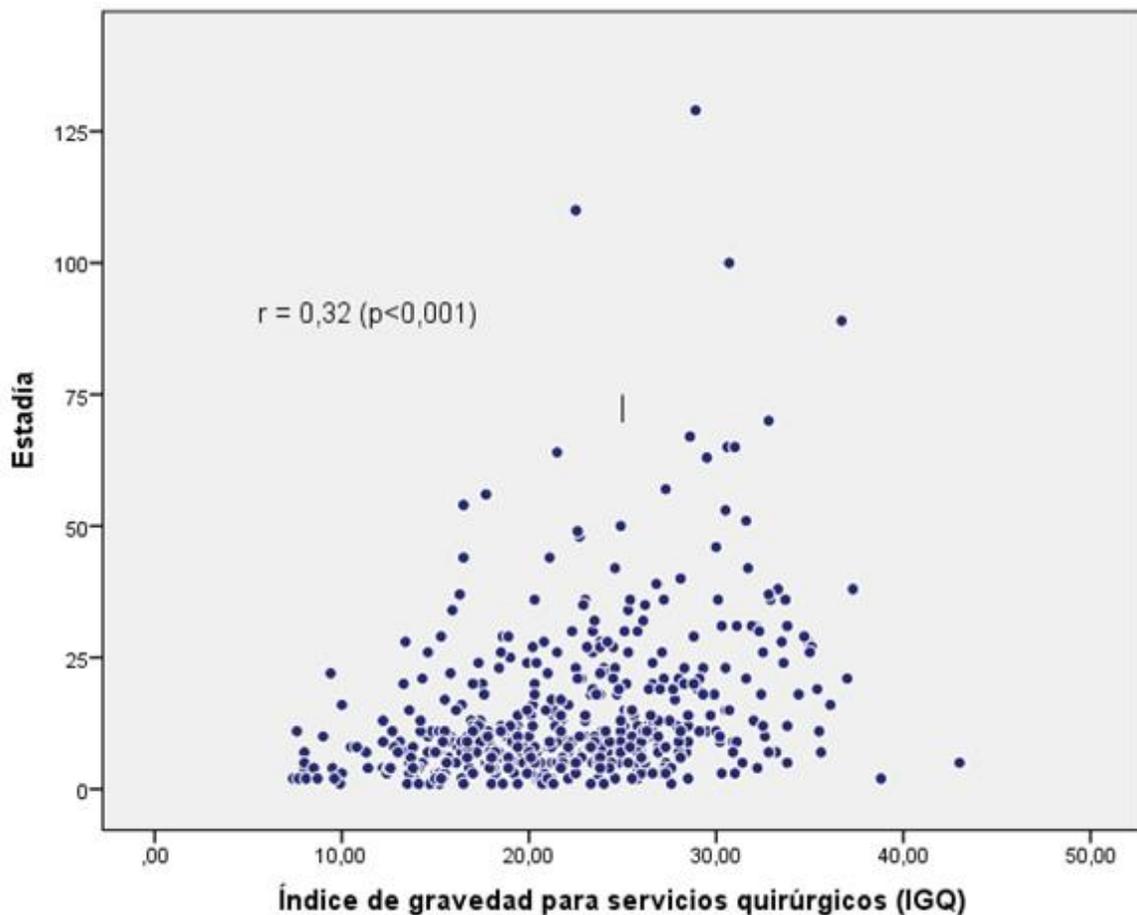


Figura 1. Diagrama de dispersión entre el índice de gravedad para servicios quirúrgicos cuantitativo y la estadía.

La asociación entre el índice en su forma ordinal y el estado al egreso se muestra en la Tabla 3. En la muestra estudiada se encontraron 27 fallecidos los cuales 25 (92,6%) clasificaron en el nivel de gravedad mayor para la escala

ordinal del índice de gravedad. La asociación se mostró significativa en general ($p < 0,001$) y el valor del coeficiente τC (0,15) indica una fuerte asociación.

Niveles del índice	Estado al egreso				Estadísticos y valor de p*
	Vivo		Fallecido		
	Número	%	Número	%	
1	124	25,6	0	0,0	$\chi^2=71,707$ $p < 0,001$ $\tau C=0,148$
2	260	53,7	2	7,4	
3	100	20,7	25	92,6	
Total (511)	484	100	27	100	

*Prueba de asociación entre nivel de gravedad y estado al egreso.

Tabla 3. Distribución de pacientes según niveles del índice de gravedad para servicios quirúrgicos ordinal y estado al egreso.

Validez de criterio

La asociación entre el índice de gravedad para servicios quirúrgicos y el índice de gravedad de Horn se observa en

la Figura 2. El coeficiente de correlación de Pearson entre ambos índices alcanzó un valor de 0,722 (intervalo de confianza de 95%: 0,677-0,761).

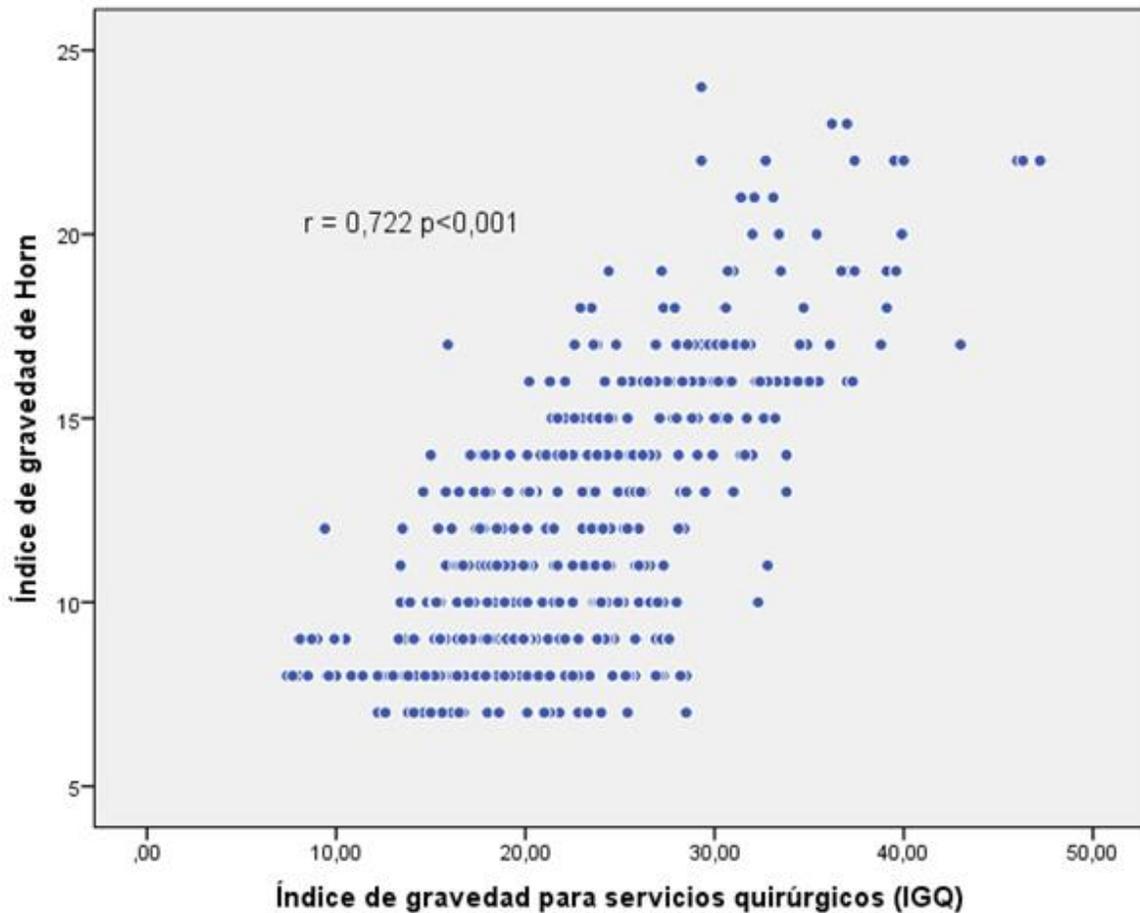


Figura 2. Diagrama de dispersión entre Índice de gravedad quirúrgico e Índice de gravedad de Horn.

De igual forma la correlación entre el índice propuesto y el riesgo de muerte por la escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* es 0,454 (intervalo de confianza de 95%: 0,388-0,514). Con el riesgo de morbilidad según este mismo sistema, la correlación es 0,539 (intervalo de confianza de 95%: 0,462-0,607).

Consistencia interna

En relación con la consistencia interna, el coeficiente α de Cronbach fue de 0,153 que con los ítems estandarizados subió a 0,791. Con respecto a la correlación entre ítems, el valor más alto (0,79) se observó entre el uso de procedimientos urgentes y la aparición de algún fallo de órgano. Entre la variable "etiología de las enfermedades asociadas" y la variable "localización de las enfermedades asociadas" la

correlación fue de 0,77. Con valores del coeficiente de correlación entre 0,6 y 0,7 se encontraron los pares de variables: estancia en unidad de cuidados intensivos y aparición de complicaciones generales (0,662); aparición de algún fallo de órgano y aparición de alguna complicación general (0,635); estancia en unidad de cuidados intensivos y necesidad de algún proceder urgente (0,617); aparición de algún fallo de órgano y estancia en unidad de cuidados intensivos (0,687).

Otros elementos de la consistencia interna se observan en la Tabla 4. La eliminación de cualquiera de los ítems excepto la edad provocaría una disminución importante en el α de Cronbach, algo que le confiere importancia a los ítems en cuanto a consistencia interna se refiere.

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	α de Cronbach si se elimina el elemento
Edad	8,47	12,26	0,298	0,188	0,682
Etiología del diagnóstico principal	60,75	271,67	0,286	0,480	0,139
Etiología de las enfermedades asociadas	61,39	269,81	0,357	0,620	0,132
Localización de enfermedades asociadas	60,93	265,52	0,363	0,628	0,119
Intervención	59,69	269,58	0,134	0,334	0,136
Tipo de intervención	61,95	276,26	0,01	0,232	0,154
Complicaciones de la herida quirúrgica	62,07	275,18	0,137	0,278	0,15
Complicaciones generales	62,04	274,61	0,168	0,558	0,148
Re intervención	62,11	275,67	0,111	0,415	0,151
Procederes invasivos	61,52	274,82	0,095	0,193	0,149
Procederes urgentes	62,11	275,21	0,176	0,657	0,15
Estancia en UCI	62,05	274,90	0,151	0,617	0,149
Fallo de órgano	62,09	275,03	0,172	0,721	0,149
Efectos residuales	61,08	267,85	0,292	0,567	0,127
Transfusiones	62,04	273,94	0,233	0,274	0,146

UCI: unidad de cuidados intensivos.

Tabla 4. Elementos de la consistencia interna para el índice de gravedad para servicios quirúrgicos.

Concordancia inter observadores (confiabilidad)

El valor obtenido del coeficiente de confiabilidad (R) fue de 0,98 lo cual indica que el 98% de la varianza total de las mediciones, se debió a los pacientes.

Con el índice de gravedad para servicios quirúrgicos ordinal calculamos el coeficiente κ global ponderado, que mide la

concordancia entre evaluadores. Se obtuvo un coeficiente de κ ponderado global de 0,87, la categoría 3 tuvo el mayor coeficiente (0,88) y el menor fue de 0,63 para la categoría 2. Todos fueron significativamente diferentes de cero ($p < 0,05$) (Tabla 5). Sin embargo, los intervalos de confianza señalan que κ para las categorías 1 y 2 podría tomar valores menores de 0,5.

Categorías	κ	IC (95 %)	p
Categoría 1	0,68	0,35-1,00	0,0022
Categoría 2	0,63	0,24-1,00	0,0048
Categoría 3	0,88	0,64-1,00	0,0001
κ global	0,87	0,74-1,00	0,0001

IC: intervalo de confianza.

Tabla 5. Valor de coeficiente κ

Discusión

En la evaluación del desempeño hospitalario a través de indicadores de resultados, los índices que miden gravedad de la enfermedad son de gran utilidad. Buena parte de este ajuste se realiza utilizando índices que engloban y cuantifican la gravedad de cada paciente durante el período de hospitalización. Así permiten deslindar la parte de los resultados que puede deberse a la calidad de la atención recibida. En el presente trabajo se presentan los resultados de la reevaluación de la validez y confiabilidad de un índice de gravedad de carácter general para servicios quirúrgicos, desarrollado y validado en la década de los noventa, en el Hospital Hermanos Ameijeiras de La Habana, Cuba.

La evaluación de la validez conceptual (más conocida como validez de constructo) muestra aristas complicadas. Se trata de encontrar aspectos indirectamente relacionados con lo que intenta medir el índice, como son la estadía y el estado al egreso. En cuanto el estado al egreso se puede pensar que, si el índice propuesto realmente mide el nivel de gravedad de un paciente como los que se incluyen en este estudio, entonces el índice como variable debería mostrar niveles más altos para los fallecidos que para los egresados vivos. También supuestamente el índice debería asociarse positivamente con la estadía en pacientes egresados vivos. Dados los resultados del estudio, la validez conceptual se considera aceptable, pues corrobora nuestras hipótesis tanto en la versión ordinal como en la cuantitativa. Encontramos una situación similar en estudios previos realizados por nosotros con un índice de gravedad similar para los servicios clínicos [20]. La relación de los índices de gravedad con la mortalidad se evalúa con mucha frecuencia, en particular su capacidad para predecir mortalidad en pacientes quirúrgicos. Se trata de una relación que confirma la validez conceptual (constructo) de un índice y que además tiene una posible utilidad en el pronóstico. La escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* por ejemplo, ha sido evaluada como predictora de mortalidad postquirúrgica en varios estudios [26],[27],[28]. El índice presentado está concebido como indicador para uso en el ajuste por riesgo para evaluaciones de desempeño. Su utilización como índice pronóstico para después de la intervención quirúrgica, conllevaría ajustes que deberán evaluarse en estudios posteriores.

En relación con la estadía, aunque la asociación es significativa, el coeficiente de correlación de Pearson fue relativamente bajo. Si bien se esperaba alguna asociación directa entre el índice de gravedad para servicios quirúrgicos y la estadía, no se aspiraba a que ésta fuera alta. Se ha demostrado que la estadía no depende solamente de las características clínicas y demográficas de los pacientes, existen otros factores que pueden influir sobre el tiempo de permanencia de un paciente en el hospital como se ha enfatizado y demostrado en diferentes estudios [29],[30],[31]. Entre estos factores se incluyen desde aquellos que reflejan estado socioeconómico de los pacientes, hasta los que indican utilización de servicios hospitalarios. En el trabajo original, donde se propone por primera vez este índice, la correlación entre la estadía y la gravedad resultó aún menor. En un estudio donde se evaluaron aspectos de la validez del Índice de Gravedad Computarizado (*Computerized Severity Index, CSI*), un indicador derivado del Índice de Horn, se encontró que el 54% de la estadía podía explicarse por la gravedad cuando esta se unía a los Grupos de Diagnósticos Relacionados [32]. Otros autores sin embargo, encuentran que la gravedad incluso conjuntamente con otras variables es capaz de explicar menos del 30% de la estadía [33].

La validez de criterio suele ser cómoda de evaluar cuando se cuenta con otra medida que puede considerarse patrón. En nuestro caso empleamos dos escalas para evaluar validez de criterio: el *Severity of Illness Index* propuesto por Horn y colaboradores, y la escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity*. La escala de Horn mostró validez y confiabilidad aceptables en el estudio de validación realizado después de su aparición. La escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* ha sido validada en diversos estudios, principalmente demostrando su capacidad para predecir mortalidad o morbilidad en pacientes postquirúrgicos y para realizar ajuste de riesgo en la evaluación de resultados en el área de la cirugía [34].

La alta correlación entre el Índice de Gravedad de la Enfermedad de Horn y el índice de gravedad cuantitativo fue similar a la encontrada en la primera validación realizada al índice de gravedad para servicios quirúrgicos. Esta alta correlación podría sin dudas 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 al de la validación y

en un periodo de tiempo anterior. No obstante, lograr un indicador que tenga alta correlación con el Índice de Horn resalta la validez del índice propuesto, sobre todo, si como sucede aquí, el nuevo indicador es de aplicación menos complicada.

Con el riesgo de mortalidad y de morbilidad obtenidos a través de la escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* nuestro índice mostró una correlación significativa, pero no muy alta (alrededor de 0,5). En general, la validez de criterio (concurrente, como en este caso), solo demuestra que el nuevo instrumento no está muy alejado de lo que, respecto al concepto que se mide, han alcanzado otros instrumentos ya consolidados. La escala *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* incluye 14 variables fisiológicas (no siempre medidas en un paciente quirúrgico) y seis variables que dan cuenta de la gravedad de la intervención. Resulta sin dudas una escala más compleja que la que se propone en este trabajo. En esos términos, las asociaciones o correlaciones alcanzadas por el índice de gravedad para servicios quirúrgicos, se pueden dar por aceptables. Una revisión sistemática reciente donde se evalúan propiedades de escalas endoscópicas que miden actividad de la enfermedad de Crohn encuentra coeficientes de correlación para validez de criterio que oscilan entre 0,4 y 0,6 [35].

En el terreno de la confiabilidad, la concordancia entre jueces resulta la característica más deseable para la utilidad que se espera del índice que se propone. El índice apropiado para medir la concordancia entre jueces (o evaluadores) para variables continuas, es el coeficiente de correlación intraclase o coeficiente de confiabilidad (R). Cuando un atributo es evaluado en distintos sujetos mediante una escala cuantitativa aplicada por más de un observador, este coeficiente indica qué parte de la varianza total de las observaciones obtenidas se debe a los sujetos. Un coeficiente alto indica que la variación entre evaluadores y otras fuentes es pequeña y, por tanto, la escala de medición resulta confiable.

Lee y colaboradores, por ejemplo, sugieren que para que dos mediciones se puedan considerar intercambiables, el intervalo de confianza estimado para este coeficiente debía tener un límite inferior por encima de 0,75 [36]. Sin embargo Müller y Büttner critican esta regla, aduciendo que el coeficiente de correlación intraclase depende de la varianza de la población y que no tiene un real significado [37]. A pesar de que un coeficiente de correlación intraclase mayor de 0,9 como el encontrado aquí, parece a todas luces aceptable, o quizás muy aceptable, la mayor parte de los textos o documentos que tratan sobre este coeficiente no expresan el valor que este debe tener para ser considerado como tal. Probablemente, porque este coeficiente adopta diferentes formas y porque un valor umbral que sea independiente del contexto donde fue obtenido es imposible y quizás indeseable.

Otro coeficiente empleado para evaluar concordancia entre jueces o evaluadores es el coeficiente κ de Cohen,

propuesto en 1960 y luego perfeccionado por Fleiss y Cohen [38]. Este coeficiente está concebido para variables cualitativas con dos o más categorías y ha sido ampliamente empleado en el campo de la salud. En nuestro caso, se utilizó gracias a que estamos proponiendo dos versiones del índice de gravedad para servicios quirúrgicos, una de las cuales contempla tres categorías ordinales. Para esta versión nuestro κ global fue de 0,87 (intervalo de confianza 95%: 0,74-1,00). Landis y Koch, secundados por Fleiss sugieren que valores por encima de 0,7 pueden considerarse muy aceptables [39]. En cuanto a la concordancia por categorías, lo relevante es que hubo una mayor concordancia en la clasificación de los pacientes en la categoría 3. Es posible que sea más sencillo clasificar a los sujetos más graves y esto se refleje en la concordancia entre evaluadores. Por otro lado, puede ser que la mayor parte de las veces lo que más interese sea no cometer errores en la clasificación de los pacientes más graves, y esto resulte una fortaleza del índice propuesto. Los intervalos de confianza señalan que si bien el κ global solo puede tomar valores mayores de 0,7 (con 95% de confianza), la concordancia para las categorías 1 y 2 podría ser realmente baja.

La consistencia interna del índice de gravedad para servicios quirúrgicos aplicado al servicio de cirugía general se calificaría de aceptable, si tenemos en cuenta que Bland y Altman señalan que un coeficiente α de Cronbach entre 0,7 y 0,8 es satisfactorio, cuando la escala se utiliza para comparar grupos y esto es similar al uso que se le daría en el ajuste por riesgo [40].

Streiner y colaboradores [41] señalan que en algunas escalas los ítems son expresiones del efecto que tiene el concepto que se evalúa y ponen el ejemplo de la ansiedad que suele medirse a través de sus manifestaciones. En tanto, en otras escalas el concepto que se mide está conformado por la unión de sus ítems y se ilustra con calidad de vida. En el primer caso puede ser deseable que los acápites estén muy correlacionados entre sí y tengan alta "consistencia interna". Pero en el segundo, podría ésta no ser una característica importante. El índice de gravedad para servicios quirúrgicos, en nuestra opinión, pertenece al segundo caso ya que la "gravedad de la enfermedad", es el resultado de la acción conjunta de diferentes variables. Lo que se espera es que la adición de todas ellas, conformen una medida real de la gravedad del enfermo en términos generales. Algo similar señala Feinstein en su libro *Clinimetrics* [42], donde afirma que "la evaluación de la consistencia interna se ha utilizado poco debido a que muchos índices se construyen de forma libre y los ítems que sirven para el cálculo de ellos generalmente tienen papeles e importancia diferentes".

En cuanto a los ítems que contiene el índice evaluado, pueden señalarse inconvenientes que constituyen limitaciones de la escala. Algunos puntos son difíciles de evaluar. El más ilustrativo de esto es la intervención o grupo de intervención, sobre todo porque generalizarlo a otros contextos resulta complicado. Otros ítems como el uso de procedimientos urgentes o invasivos podrían tener un alto grado de subjetividad, en particular para personal

administrativo no especializado. Esto podría paliarse con entrenamiento a evaluadores, pero se trata de algo que con esta escala se pretende minimizar. Sin embargo, la introducción de procesos automatizados en la confección de la historia clínica o buena parte de ella, permitiría que los profesionales que participan directamente en el proceso diagnóstico y terapéutico introdujeran directamente los datos en formularios digitales debidamente adaptados a las condiciones de cada lugar. En tal caso, las limitaciones mencionadas antes se verían minimizadas.

Otra posible limitación es que este estudio solo se hizo en un servicio de cirugía general y este índice está concebido para servicios de cualquier especialidad quirúrgica. Podría ser necesaria su validación en servicios quirúrgicos con características especiales como oftalmología u otorrinolaringología.

Por último, es necesario enfatizar que los datos para ajuste de riesgo continúan siendo una preocupación de las autoridades sanitarias que deben evaluar el desempeño de los centros de atención médica, particularmente de los hospitales. En un artículo reciente publicado por la revista Cirugía Española, se señalan las deficiencias que aún persisten en las bases de datos administrativas en cuanto a datos disponibles para realizar ajuste por riesgo [43].

La mayor parte de los índices de gravedad conocidos son específicos para enfermedades y no son aptos para el ajuste por riesgo. Su utilidad se expresa sobre todo en el campo del pronóstico o predicción de futuros desenlaces en pacientes individuales. El ajuste por riesgo en el mundo se realiza principalmente a través del sistema de Grupos de Diagnósticos Relacionados, *Diagnostic Related Groups* (o derivados de éste). No obstante este sistema, además de haber sido criticado por no ser capaz de recoger toda la gravedad de los pacientes, es un sistema comercial con alto precio en el mercado. Su uso principal se relaciona con la contención de los costos hospitalarios mediante pagos prospectivos y otros sistemas de reembolso por atenciones sanitarias en hospitales donde, supuestamente, se incentiva también la calidad de las prestaciones [44].

El índice revalidado en este estudio es mucho más sencillo de aplicar y su publicación en una revista de acceso abierto garantiza su amplia divulgación. Por tanto podrá ser utilizado por cualquier centro hospitalario que, como es necesario, realice comparaciones de desempeño de su servicio de cirugía general en el espacio, en el tiempo o incluso para comparar desempeño de equipos quirúrgicos o cirujanos en particular, a través de cualquier indicador de resultados. A pesar de que, en el fondo se trata de una escala que puede tratarse como tecnología, la intención es que pueda ser utilizado por cualquier servicio o centro que lo necesite introduciéndole incluso adaptaciones que permitan su uso en diferentes escenarios y situaciones.

El índice (escala o sistema) *Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* [19] utilizado en este estudio como estándar de oro para evaluar la validez de criterio, tiene varias similitudes con el índice que presentamos. Sin embargo,

requiere datos de variables bioquímicas y hematológicas que pueden dificultar su uso en escenarios donde no se indiquen este tipo de exámenes en todos los pacientes. No obstante, este índice y su modificación posterior, el *Portsmouth Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity* [45], han sido utilizados y validados en diversos escenarios. Una comparación empírica con el que presentamos aquí, permitiría fundamentar apropiadamente la selección de uno u otro basado en la relación costo/efectividad.

Conclusiones

Los resultados obtenidos en la validación del índice de gravedad indican que conserva sus propiedades psicométricas y que puede, por lo tanto, ser empleado en evaluaciones de desempeño del servicio de cirugía general que conlleven comparaciones en tiempo o espacio. De esta forma, contribuye a enriquecer el número de instrumentos que pueden utilizarse en el ajuste por riesgo para las evaluaciones periódicas de desempeño de estos servicios.

Por su relativa sencillez para profesionales de la salud, es recomendable su uso en escenarios donde pueda automatizarse la entrada de datos y, de ser posible, la hagan directamente los profesionales involucrados en la atención a pacientes: médicos residentes, internos o enfermeras especializadas en áreas quirúrgicas.

Notas

Aspectos éticos

La Revista tiene constancia de que el comité ético científico del Hospital Hermanos Ameijeiras de La Habana, Cuba, tuvo conocimiento sobre este estudio y su posible publicación en una revista de difusión biomédica.

Declaración de conflictos de intereses

Los autores han completado el formulario de declaración de conflictos de intereses del ICMJE, y declaran no haber recibido financiamiento para la realización del reporte; no tener relaciones financieras con organizaciones que podrían tener intereses en el artículo publicado, en los últimos tres años; y no tener otras relaciones o actividades que podrían influir sobre el artículo publicado. Los formularios pueden ser solicitados contactando al autor responsable o a la dirección editorial de la Revista.

Financiamiento

Los autores declaran que no hubo fuentes de financiación externas.

Referencias

1. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care*. 2003 Dec;15(6):523-30. | [PubMed](#) |
2. Bowling A. Evaluating health services: multidisciplinary collaboration. *Research methods in health, Investigating health and health services*. 2nd ed. Philadelphia: Open University Press; 2002.

3. Iezzoni LI. Risk adjustment for performance measurement. Experiences, Challenges and Prospects. En: Performance Measurement for Health System Improvement. Cambridge: Cambridge University Press; 2009:251-85.
4. Palma-Ruiz M, Sendra-Gutiérrez JM, Sarría-Santamera A. Sistemas de ajuste de riesgo en evaluación de servicios de salud. Madrid; 2007. | [Link](#) |
5. Iezzoni LI. 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. *Ann Intern Med.* 1996 Jun 15;124(12):1079-85. | [PubMed](#) |
6. Iezzoni LI. Risk Adjustment for Measuring Healthcare Outcomes. 4th ed. Chicago: Health Administration Press; 2013.
7. DesHarnais SI, Forthman MT, Homa-Lowry JM, Wooster LD. Risk-adjusted clinical quality indicators: indices for measuring and monitoring rates of mortality, complications, and readmissions. *Qual Manag Health Care.* 2000 Fall;9(1):14-22. | [PubMed](#) |
8. Ding YY. Risk adjustment: towards achieving meaningful comparison of health outcomes in the real world. *Ann Acad Med Singapore.* 2009 Jun;38(6):552-7. | [PubMed](#) |
9. Quinn K. After the revolution: DRGs at age 30. *Ann Intern Med.* 2014 Mar 18;160(6):426-9. | [CrossRef](#) | [PubMed](#) |
10. Gonnella JS, Louis DZ, Gozum MVE, Callahan CA, Barnes CA, editors. Disease Staging Clinical And Coded Criteria. Version 5.26. Ann Arbor: Thomson Reuters; 2009.
11. Cots F, Castells X, Mercadé L, Torre P, Riu M. Risk adjustment: beyond patient classification systems. *Gac Sanit.* 2001 Sep-Oct;15(5):423-31. | [PubMed](#) |
12. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol.* 2003 Mar;56(3):221-9. | [PubMed](#) |
13. Salluh JI, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care.* 2014 Oct;20(5):557-65. | [CrossRef](#) | [PubMed](#) |
14. Horn SD, Horn RA, Sharkey PD. The Severity of Illness Index as a severity adjustment to diagnosis-related groups. *Health Care Financ Rev.* 1984;Suppl:33-45. | [PubMed](#) |
15. Horn SD, Horn RA. Reliability and validity of the Severity of Illness Index. *Med Care.* 1986 Feb;24(2):159-78. | [PubMed](#) |
16. Hall BL, Bilimoria KY, Ko CY. Investigations using clinical data registries: observational studies and risk adjustment. *Surgery.* 2009 Jun;145(6):602-10. | [PubMed](#) |
17. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998 Jan;36(1):8-27. | [PubMed](#) |
18. Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res.* 2011 Jun 10;11:146. | [CrossRef](#) | [PubMed](#) |
19. Copeland GP. The POSSUM system of surgical audit. *Arch Surg.* 2002 Jan;137(1):15-9. | [PubMed](#) |
20. Jiménez Paneque RE, Vázquez García J, Fariñas Seijas H. [Construction and validation of a severity of illness index of patients hospitalized in clinical areas]. *Gac Sanit.* 1997 May-Jun;11(3):122-30. | [PubMed](#) |
21. Jiménez Paneque RE, Domínguez Alonso E, Fariñas Seijas H, Fuentes Valdés E. Construcción y validación de un índice de gravedad para pacientes hospitalizados en áreas quirúrgicas. *Rev Cubana Salud Pública.* 1999: 154-66. | [Link](#) |
22. Jiménez R, López L, Dominguez D, Fariñas H. Difference between observed and predicted length of stay as an indicator of inpatient care inefficiency. *Int J Qual Health Care.* 1999 Oct;11(5):375-84. | [PubMed](#) |
23. Tamargo Barbeito TO, Jiménez Paneque RE, Álvarez Rodríguez JM, Terry Villa O. Utilidad del Índice de Gravedad de Servicios Clínicos del Hospital Hermanos Ameijeiras. *Rev Cuba Salud Pública.* 1999, Editorial Ciencias Médicas; 2013;39(1):4-18. | [Link](#) |
24. Horn SD. Measuring severity of illness: comparisons across institutions. *Am J Public Health.* 1983 Jan;73(1):25-31. | [PubMed](#) |
25. Campillo-Soto A, Flores-Pastor B, Soria-Aledo V, Candel-Arenas M, Andrés-García B, Martín-Lorenzo JG, Aguayo-Albasini JL. [The POSSUM scoring system: an instrument for measuring quality in surgical patients]. *Cir Esp.* 2006 Dec;80(6):395-9. | [PubMed](#) |
26. Hong S, Wang S, Xu G, Liu J. Evaluation of the POSSUM, p-POSSUM, o-POSSUM, and APACHE II scoring systems in predicting postoperative mortality and morbidity in gastric cancer patients. *Asian J Surg.* 2015 Sep 23. pii: S1015-9584(15)00089-5. | [CrossRef](#) | [PubMed](#) |
27. Tran Ba Loc P, du Montcel ST, Duron JJ, Levard H, Suc B, Descottes B, et al. Elderly POSSUM, a dedicated score for prediction of mortality and morbidity after major colorectal surgery in older patients. *Br J Surg.* 2010 Mar;97(3):396-403. | [CrossRef](#) | [PubMed](#) |
28. Elías AC, Matsuo T, Grion CM, Cardoso LT, Verri PH. [POSSUM scoring system for predicting mortality in surgical patients]. *Rev Esc Enferm USP.* 2009 Mar;43(1):23-9. | [PubMed](#) |
29. Tan SS, Hakkaart-van Roijen L, van Ineveld BM, Redekop WK. Explaining length of stay variation of episodes of care in the Netherlands. *Eur J Health Econ.* 2013 Dec;14(6):919-27. | [CrossRef](#) | [PubMed](#) |
30. Moore L, Cisse B, Batomen Kuimi BL, Stelfox HT, Turgeon AF, Lauzier F, Clément J, Bourgeois G. Impact of socio-economic status on hospital length of stay following injury: a multicenter cohort study. *BMC Health Serv Res.* 2015 Jul 25;15:285. | [CrossRef](#) | [PubMed](#) |
31. Budäus L, Morgan M, Abdollah F, Zorn KC, Sun M, Johal R, et al. Impact of annual surgical volume on length of stay in patients undergoing minimally invasive prostatectomy: a population-based study. *Eur J Surg Oncol.* 2011 May;37(5):429-34. | [CrossRef](#) | [PubMed](#) |
32. Horn SD, Sharkey PD, Buckle JM, Backofen JE, Averill RF, Horn RA. The relationship between severity of illness and hospital length of stay and mortality. *Med Care.* 1991 Apr;29(4):305-17. | [PubMed](#) |
33. Sahadevan S, Earnest A, Koh YL, Lee KM, Soh CH, Ding YY. Improving the diagnosis related grouping model's ability to explain length of stay of elderly medical inpatients by incorporating function-linked variables.

- Ann Acad Med Singapore. 2004 Sep;33(5):614-22. | [PubMed](#) |
34. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg. 1991 Mar;78(3):355-60. | [PubMed](#) |
35. Khanna R, Nelson SA, Feagan BG, D'Haens G, Sandborn WJ, Zou GY, et al. Endoscopic scoring indices for evaluation of disease activity in Crohn's disease. Cochrane Database Syst Rev. 2016 Aug 8;(8):CD010642. | [CrossRef](#) | [PubMed](#) |
36. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. Comput Biol Med. 1989;19(1):61-70. | [PubMed](#) |
37. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. Stat Med. 1994;13:2465-76.
38. Fleiss JL, Levin B, Paik MC. The Measurement of Interrater Agreement. Statistical Methods for Rates and Proportions. John Wiley & Sons;2003:598-623.
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74. | [PubMed](#) |
40. Bland JM, Altman DG. Cronbach's alpha. BMJ. 1997 Feb 22;314(7080):572. | [PubMed](#) |
41. Streiner DL, Norman GR, Cairney J. Health Measurement Scales. A practical guide to their development and use. Oxford: Oxford University Press; 2015:93.
42. Feinstein AR. Clinimetrics. New Haven: Yale University Press; 1987: xi, 272.
43. Rodrigo-Rincón I, Martín-Vizcaíno MP, Tirapu-León B, Zabalza-López P, Abad-Vicente FJ, Merino-Peralta A, et al. Validez de las bases de datos administrativas para realizar ajustes de riesgo en el análisis de los efectos adversos producidos en pacientes quirúrgicos. Cirugía Española. 2016 Mar;94(3):165-74. | [Link](#) |
44. Goldfield N. The Evolution of Diagnosis-Related Groups (DRGs). Qual Manag Health Care. 2010 Jan;19(1):3-16. | [Link](#) |
45. Prytherch R, Whiteley MS, Higgins B, Weaver PC, Prout GW, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Br J Surg. 1998 Sep;85(9):1217-20. | [CrossRef](#) |

Correspondencia a:

[1] Calle San Lázaro 701
 Centro Habana
 La Habana
 Cuba



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.