



ORIGINAL

Validación clínica de 2 agrupadores de morbilidad en el ámbito de atención primaria



Montse Clèries^{a,*}, David Monterde^b, Emili Vela^a, Àlex Guarga^c, Luis García Eroles^d, Pol Pérez Sust^e y Grupo de validación[◇]

^a Unidad de Información y Conocimiento, Servicio Catalán de la Salud, Departamento de Salud de la Generalidad de Cataluña, Barcelona, España

^b Oficina de Estadística, Sistemas de Información, Instituto Catalán de la Salud, Departamento de Salud de la Generalidad de Cataluña, Barcelona, España

^c Área de Atención Sanitaria, Servicio Catalán de la Salud, Departamento de Salud de la Generalidad de Cataluña, Barcelona, España

^d Gerente de Sistemas de Información, Servicio Catalán de la Salud, Departamento de Salud de la Generalidad de Cataluña, Barcelona, España

^e Coordinador general de las tecnologías de la información y comunicación, Departamento de Salud de la Generalidad de Cataluña, Barcelona, España

Recibido el 23 de marzo de 2018; aceptado el 25 de septiembre de 2018

Disponible en Internet el 12 de febrero de 2019

PALABRAS CLAVE

Agrupador de morbilidad;
Multimorbilidad;
Validación clínica

Resumen

Introducción: Los grupos de morbilidad ajustados (GMA) y los clinical risk groups (CRG) son herramientas de estratificación poblacional basada en la morbilidad que permiten clasificar a los pacientes en categorías mutuamente excluyentes.

Objetivo: Comparar la estratificación, según niveles de complejidad, proporcionada por los GMA con la de los CRG y con la realizada ad-hoc por los evaluadores.

Diseño: Muestra aleatoria por estratos de riesgo de morbilidad.

Emplazamiento: Cataluña.

Participantes: Cuarenta médicos de atención primaria emparejados 2 a 2.

Intervenciones: Cada pareja de evaluadores tuvo que validar 25 historias.

Mediciones principales: Se evaluó la concordancia entre evaluadores, y entre los evaluadores y los resultados obtenidos por los 2 agrupadores de morbilidad con el índice kappa, la sensibilidad, la especificidad, el valor predictivo positivo y el negativo.

Resultados: La concordancia entre evaluadores se situó alrededor del valor kappa 0,75 (valor medio = 0,67), entre los GMA y los evaluadores fue similar (valor medio = 0,63), y más elevada que para los CRG (valor medio = 0,35). Los profesionales otorgaron una puntuación de 7,5 a la bondad de ambos agrupadores, aunque para los estratos de mayor complejidad, según asignación de los profesionales, los GMA obtuvieron mejores puntuaciones que los CRG. Los

* Autor para correspondencia.

Correo electrónico: mcleries@catsalut.cat (M. Clèries).

◇ Los nombres de los componentes del grupo de validación están relacionados en el [anexo A](#).

profesionales prefirieron mayoritariamente los GMA frente a los CRG. Estas diferencias se incrementaron con el aumento de la complejidad de los pacientes según el criterio clínico.

Globalmente, se encontró menos de un 2% de errores graves de clasificación proporcionada por ambos agrupadores.

Conclusión: Los profesionales consideraron que ambos agrupadores clasificaban adecuadamente a la población, aunque los GMA tienen un mejor comportamiento en los estratos que los profesionales identifican como de mayor complejidad. Además, en la mayoría de los casos, los evaluadores clínicos prefirieron los GMA.

© 2019 El Autor. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Morbidity grouper;
Multimorbidity;
Clinical Validation

Clinical validation of 2 morbidity groups in the primary care setting

Abstract

Introduction: Adjusted Morbidity Groups (GMAs) and the Clinical Risk Groups (CRGs) are population morbidity based stratification tools which classify patients into mutually exclusive categories.

Objective: To compare the stratification provided by the GMAs, CRGs and that carried out by the evaluators according to the levels of complexity.

Design: Random sample stratified by morbidity risk.

Location: Catalonia.

Participants: Forty paired general practitioners in the primary care, matched pairs.

Interventions: Each pair of evaluators had to review 25 clinical records.

Main outputs: The concordance by evaluators, and between the evaluators and the results obtained by the 2 morbidity tools were evaluated according to the kappa index, sensitivity, specificity, and positive and negative predicted values.

Results: The concordance between general practitioners pairs was around the kappa value 0.75 (mean value = 0.67), between the GMA and the evaluators was similar (mean value = 0.63), and higher than for the CRG (mean value = 0.35). The general practitioners gave a score of 7.5 over 10 to both tools, although for the most complex strata, according to the professionals' assignment, the GMA obtained better scores than the CRGs. The professionals preferred the GMAs over the CRGs. These differences increased with the complexity level of the patients according to clinical criteria.

Overall, less than 2% of serious classification errors were found by both groupers.

Conclusion: The evaluators considered that both grouping systems classified the studied population satisfactorily, although the GMAs showed a better performance for more complex strata. In addition, the clinical raters preferred the GMAs in most cases.

© 2019 The Author. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

Los grupos de morbilidad ajustados (GMA) y los clinical risk groups (CRG) son herramientas de estratificación poblacional que, a partir de los problemas de salud disponibles en los sistemas de información administrativos, permiten clasificar a la población en diferentes categorías mutuamente excluyentes^{1,2}. Cabe destacar 2 diferencias importantes entre ambas clasificaciones: por un lado, los GMA se centran en la complejidad de la multimorbilidad, mientras que los CRG se focalizan en el gasto; y por otro lado, la variable «peso relativo» para los GMA es individual, y en los CRG es grupal, propia de cada categoría específica de CRG (1.000 categorías, aproximadamente).

La multimorbilidad es la norma y no la excepción³, cosa que preocupa a los profesionales de atención primaria (AP)⁴, y existe cierta dificultad para especificarla^{5,6}. También hay

dificultad para definir su complejidad y los factores o condiciones determinantes^{7,8}.

Garantizar la fiabilidad y la concordancia de los instrumentos de medida son aspectos fundamentales en las ciencias de la salud⁹. Para medir la reproducibilidad, la concordancia interevaluador es importante en las validaciones clínicas¹⁰. El desarrollo de los GMA se llevó a cabo a partir de algoritmos estadísticos que han mostrado muy buenos resultados a nivel poblacional¹, pero era necesaria una validación clínica para conocer la valoración de los profesionales sanitarios. La validación de estas herramientas se orientó, especialmente, al ámbito de AP para poder evaluar su utilidad. Actualmente, se dispone de la información que proporcionan los GMA en la mayoría de las historias clínicas electrónicas de AP ?eCAP? de Cataluña.

Los objetivos son: i) comparar las clasificaciones proporcionadas por los GMA y los CRG; ii) comparar la

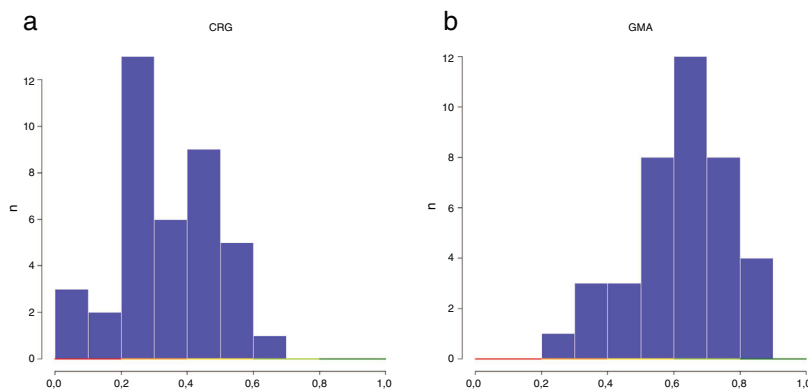


Figura 1 Distribución del índice kappa en los profesionales al analizar la concordancia de clasificación, según su criterio y el de los 2 agrupadores. La banda de color indica la interpretación del índice kappa según Landis y Koch.

estratificación de los GMA y los CRG con la realizada ad-hoc por médicos de AP, y iii) encontrar posibles áreas de mejora de los GMA.

Material y métodos

Muestreo

Se estratificó con CRG y GMA la población de Cataluña que durante el año 2013 utilizó los servicios sanitarios (6.100.360 personas). Como las clasificaciones que proporcionan los CRG y los GMA no son idénticas, se optó por generar pirámides de riesgo, agrupando en 5 estratos (percentiles de riesgo 50, 85, 95 y 99). Se realizó una muestra aleatoria de 500 casos focalizada en las discrepancias de clasificación entre los agrupadores de morbilidad disponibles (GMA y CRG). Se emparejaron los evaluadores y cada uno analizó 25 casos, uno de cada casilla de la tabla obtenida al cruzar la estratificación proporcionada por ambas herramientas (T-1 Anexo B). De los 42 evaluadores iniciales, 2 se retiraron (resultando 19 parejas de evaluadores que revisaron simultáneamente 475 historias clínicas[HC] [950 cuestionarios completados] y 2 quedaron desaparejados, los cuales revisaron 50 HC [50 cuestionarios]).

Aunque en el diseño de la muestra se priorizó el análisis de las discrepancias para poder ahondar en las diferencias, los siguientes resultados se presentan extrapolados a la población general en un ejercicio de contextualizar mejor los resultados obtenidos. Para ello, se ponderó cada sujeto evaluado en la muestra en función del número de casos poblacionales del estrato al que pertenecía.

Evaluadores

Participaron 3 proveedores de AP con diferentes características en relación con HC (modelo propio-HP, OMI y eCAP), codificación de enfermedades (CIE-9-MC, CIAP-2, CIE-10) y diferentes experiencias en la utilización de agrupadores de morbilidad.

Se realizó un curso de formación en agrupadores de morbilidad gestionado por la Escuela de Administración Pública de Cataluña, con una parte teórica y otra práctica en la que cada profesional iniciaba la evaluación de sus casos.

Participaron 40 médicos de AP que asistieron a la formación y retornaron los formularios debidamente cumplimentados. Los formularios estaban divididos en 3 bloques: información clínica, resultados de la agrupación y las pirámides de riesgo, y preguntas relacionadas con la clasificación de los pacientes (T-2 Anexo B).

Tomando como referencia la complejidad otorgada por el evaluador, se consideró como error de clasificación del agrupador si clasificaba al paciente a más de un estrato de diferencia (T-3 Anexo B).

Métodos estadísticos

Se utilizó el índice kappa¹¹, la ponderación cuadrática que da más relevancia a las discrepancias graves, para evaluar la concordancia entre observadores y entre los resultados de los distintos agrupadores. Se analizó la sensibilidad, la especificidad, el valor predictivo positivo y el valor predictivo negativo analizando estratos dicotómicos y tomando como «gold standard» el criterio clínico. El tratamiento de datos se llevó a cabo con SPSS y R.

Resultados

En cuanto a la concordancia de la clasificación entre profesionales, la mayoría de los resultados por parejas se encontraban entre 0,1 y 0,8, concentrados en torno al valor 0,75 (valor medio = 0,67). Es decir, existía una elevada concordancia entre la mayoría de las parejas de evaluadores (Figura 1 Anexo B).

Se analizó la concordancia entre el estrato de complejidad asignado por los profesionales y el estrato obtenido por los 2 agrupadores. La concordancia entre el criterio clínico y la complejidad asignada por los CRG obtuvo un índice kappa entre 0,2 y 0,6 (valor medio = 0,35), mientras que en el caso de los GMA, los valores oscilaban entre 0,5 y 0,8 (valor medio = 0,63) (fig. 1), es decir, en los GMA la concordancia entre la complejidad asignada por el agrupador y la del evaluador era similar a la encontrada entre evaluadores, y claramente superior a la obtenida ente los CRG y el evaluador.

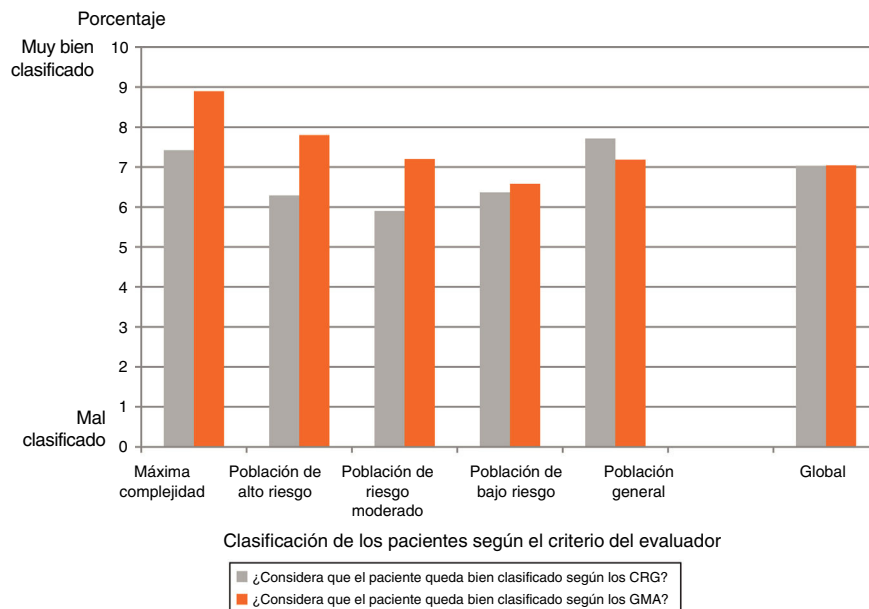


Figura 2 Bondad de las clasificaciones generadas por los agrupadores en función del nivel de complejidad otorgado por el evaluador clínico. Extrapolación al conjunto de la población.

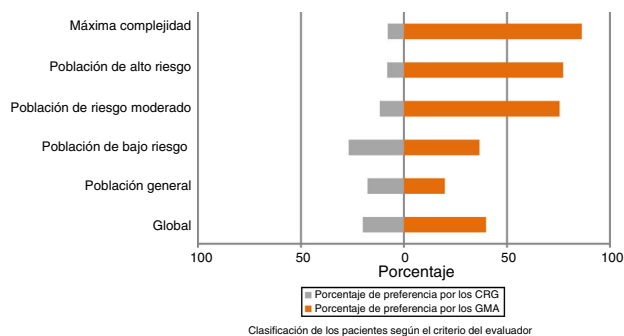


Figura 3 Preferencias de los profesionales sanitarios por uno de los agrupadores, según el nivel de complejidad otorgado por el evaluador. Extrapolación al conjunto de la población.

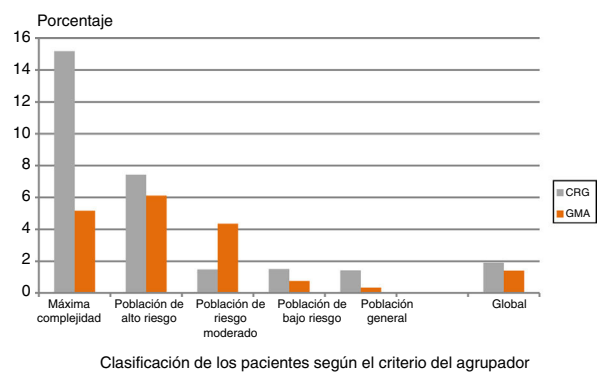


Figura 4 Porcentaje de clasificaciones erróneas según el estrato generado por el agrupador (falsos positivos). Extrapolación al conjunto de la población.

Extrapolación a la población

Analizando los datos extrapolados a la población, aunque se minimizaron las diferencias encontradas entre GMA y CRG, se mantuvieron los resultados encontrados con los datos muestrales.

Globalmente no se encontraron diferencias entre CRG y GMA, pero GMA clasifica mejor a los pacientes de niveles más complejos (fig. 2).

De la pregunta: «¿Qué clasificación considera más adecuada?», se obtuvo que, globalmente, en un 40% de los casos los clínicos preferían los GMA, en un 20% los CRG y en el 40% restante les era indiferente el agrupador. Estas diferencias se incrementaron en los niveles de mayor complejidad, según criterio clínico, llegando al 86% la preferencia por los GMA y a un 8% la predilección por los CRG (fig. 3). Mayoritariamente, cuando el agrupador preferido son los CRG, es debido a la existencia de la prescripción farmacéutica, que ofrece información complementaria a la de los diagnósticos,

mientras que en el caso de preferir los GMA, lo más valorado fueron las etiquetas clínicas que proporcionan.

Globalmente, la clasificación de los agrupadores presenta menos de un 2% de errores. En el grupo de mayor riesgo, según el agrupador, los errores (falsos positivos) en los CRG fueron del 15% y del 5% para los GMA (fig. 4). Se evaluaron también los errores de clasificación desde la perspectiva de la clasificación del evaluador (falsos negativos). Los GMA presentan un porcentaje menor de falsos negativos que los CRG (fig. 5).

Resultados sin extrapolar a la población

De las 1.000 HC revisadas, hubo errores de clasificación en el 29,5% de los CRG y en el 14,3% de los GMA. De las 475 HC emparejadas, 96 parejas de evaluadores estuvieron de acuerdo en el error de clasificación del CRG, mientras que en los GMA fueron 24.

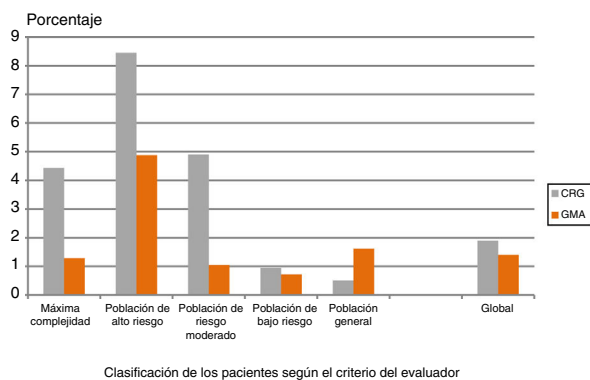


Figura 5 Porcentaje de clasificaciones erróneas según el criterio del profesional clínico (falsos negativos). Extrapolación al conjunto de la población.

Falsos positivos

Error en el estrato de máxima complejidad: para los CRG, coincidió el criterio de la pareja de evaluadores (26 parejas [52 HC] y 30 HC según criterio de un único evaluador), mientras que para los GMA, todos se basaron en el criterio de un único evaluador (12 HC). Según los evaluadores, más de la mitad de los errores de los CRG fue por sobrevalorar la información que aportan los fármacos (además de clasificar a muchos pacientes de esquizofrénicos sin serlo), mientras que los errores de los GMA, en su mayoría, fueron por sobrevalorar la enfermedad crónica, en algunos pacientes que habían hecho un bajo consumo de recursos sanitarios (posiblemente por haber fallecido en la primera parte del año o por estar ingresados en residencias) y que el clínico valoró como controlados o que habían dejado de tener riesgo. Las diferencias en el siguiente estrato, alto riesgo, fueron de un 7% en CRG y un 6% en GMA.

Falsos negativos

En el estrato de máxima complejidad según los evaluadores, en los CRG, en 12 de 43 errores coincidía el criterio de la pareja de evaluadores (6 parejas); según los evaluadores, los CRG consideraron sanos o de poca complejidad a pacientes terminales (28/43), la mayoría fallecieron o fueron ingresados en residencias el año de estudio, en algunos casos infravaloraron neoplasias activas (7/43) y diagnósticos (8/43). Por otra parte, en los de los GMA hubo 6 errores; todos afectaban solo al criterio de un único evaluador, y el problema fue la falta de diagnóstico en 3/6 casos, solo constaba información de farmacia (linfoma) o de procedimientos (respiración asistida y diálisis) no utilizada en la construcción del GMA, o por infravalorar el riesgo en algunos diagnósticos (2/6 tetraplejias y 1/6 neoplasia cerebral).

Se calculó la sensibilidad, la especificidad y los valores predictivos de los agrupadores (tabla 1); los resultados fueron muy buenos para ambos agrupadores, aunque los GMA obtuvieron mejores resultados que los CRG, sobre todo en lo que respecta al valor predictivo positivo.

Observaciones de los evaluadores

Destacan:

- La importancia de notificar correctamente todas las enfermedades.

- Tener en cuenta también otros aspectos: gravedad de las enfermedades, edad y factores sociales.
- Falta información del nivel de desarrollo de las enfermedades, hecho que puede llevar a sobrevalorar o infravalorar la complejidad.
- Pacientes con problemas mentales no quedan bien clasificados con ninguno de los agrupadores. Además, los CRG, que utilizan la prescripción farmacéutica, suelen adjudicar un diagnóstico, generalmente esquizofrenia, que no es correcto.
- Los GMA infravaloran algunos casos en que no consta el diagnóstico, y no tienen en cuenta la prescripción farmacéutica ni los procedimientos, y en algunos, los CRG sobrevaloran la complejidad por dar demasiada importancia a la prescripción.
- Se valoran muy positivamente las etiquetas que proporcionan los GMA, pero se echaron en falta algunas afecciones que podrían ayudar a valorar la complejidad (sordera, ceguera, neoplasia reciente [anterior al año de estudio], hipotiroidismo, úlceras por presión, etc.).
- En pacientes fallecidos durante el año o que han comenzado con una enfermedad importante durante el período de análisis, los CRG no detectan adecuadamente su dolencia.

Discusión

Es importante conocer la calidad de los procedimientos de medida y los errores de medida inherentes a la herramienta, entendiéndose que a mayor calidad de la medida menor magnitud de los errores y viceversa⁹. Los agrupadores analizados utilizan datos de morbilidad existentes en bases de datos clínicas o administrativas; la calidad de las mismas en relación con la precisión de la codificación, la validación y los diagnósticos secundarios va a ser fundamental¹² para obtener una buena estratificación que refleje la realidad de la población estudiada. Algunos estudios muestran que los datos administrativos son mejores que los autoinformes¹³ y que son válidos para la obtención de datos poblacionales¹⁴; otros, que existe una infranotificación de enfermedades poco relevantes¹⁵⁻¹⁸. Si la recogida de información es prospectiva, precisa y verificada, es la mejor garantía para una buena calidad de la información¹⁹. En nuestro caso, las bases de datos administrativas se notifican de manera prospectiva, alguna de ellas de manera continuada, como es el caso de la actividad urgente, y han ido mejorando gracias a su mayor utilización y a los avances tecnológicos, que permiten una mejor comunicación entre los diferentes niveles asistenciales y unas HC compartidas, que facilitan que un paciente pueda ser atendido por diferentes profesionales con el máximo de información disponible sobre él. Además, existe una validación cruzada con los sistemas de facturación.

Se conoce que para la validación de una herramienta es necesario disponer también de un buen «gold standard»^{20,21}, y en este estudio este es el criterio de los médicos de AP.

Al tratarse de una muestra con evaluadores emparejados, se ha podido analizar la reproducibilidad⁹ de las evaluaciones, que, en general, ha sido buena. Pero en relación con el análisis de los errores de clasificación según la pirámide de riesgo, no afectan por igual en todos los niveles de

Tabla 1 Sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo

Complejidad	Sensibilidad		Especificidad		Valor predictivo +		Valor predictivo –	
	CRG	GMA	CRG	GMA	CRG	GMA	CRG	GMA
1. Población de máxima complejidad	0,972	0,993	0,998	0,999	0,848	0,949	1,000	1,000
2. Población de alto riesgo	0,886	0,931	0,998	0,997	0,926	0,939	0,993	0,996
3. Población de riesgo moderado	0,945	0,988	0,989	0,994	0,985	0,957	0,992	0,998
4. Población de bajo riesgo	0,990	0,993	0,989	0,994	0,985	0,992	0,993	0,995
5. Población general	0,996	0,985	0,991	0,998	0,986	0,997	0,997	0,991

riesgo; en ambos agrupadores afectan especialmente a los pacientes clasificados como de máxima complejidad o alto riesgo y que los evaluadores no los consideran, o en aquellos que según los evaluadores sí estarían entre los pacientes más complejos y según los agrupadores no. El porcentaje de errores hallado es mayor en los CRG que en los GMA, sobre todo en los estratos de mayor riesgo. Este es un tema de gran relevancia, ya que los agrupadores han de ser capaces de proporcionar una adecuada estratificación poblacional según la complejidad de la población. Se ha encontrado un mayor grado de acuerdo intraevaluador para los errores de clasificación generados por los CRG que para los producidos por los GMA, es decir, los evaluadores mostraron menor acuerdo en los errores atribuidos a los GMA. En el ámbito de AP se utilizan diferentes metodologías para determinar pacientes lábiles, o con enfermedades específicas, como son los autocuestionarios, para realizar intervenciones preventivas o curativas²², los agrupadores actualmente se utilizan también para identificar poblaciones de pacientes crónicos complejos en los que llevar a cabo programas específicos de intervención^{1,23}, por ello disminuir el porcentaje de errores en estos estratos es de gran importancia.

Existen algunas limitaciones para determinar la complejidad de un paciente a partir únicamente de la información diagnóstica, terapéutica y de prescripción que consta en una HC; es una tarea difícil y sujeta a un cierto grado de variabilidad según la experiencia personal de cada profesional, lo que puede explicar la falta de concordancia clínica²⁴. Los agrupadores de morbilidad se elaboran básicamente a partir de los diagnósticos, que no suelen recoger la intensidad o fase de la evolución de la enfermedad, información que puede ser crucial para discriminar el estrato de morbilidad idóneo; algunos agrupadores mejoran su resultado porque utilizan otras fuentes de información como los procedimientos realizados a los pacientes o la prescripción farmacéutica. Los evaluadores manifiestan su interés por disponer de otros indicadores (sociales) que ayuden a discriminar mejor la complejidad de los pacientes, pero esta información actualmente no está disponible.

También comentan las limitaciones que presentan los agrupadores para clasificar correctamente a los pacientes con enfermedades mentales. Para solventar este problema se está trabajando en un nuevo agrupador GMA específico de problemas de salud mental y adicciones.

A partir de la validación clínica, se ha cambiado el criterio para considerar una neoplasia activa, que en el momento de la validación se refería exclusivamente a las neoplasias diagnosticadas en el año de estudio; en la versión actual se ha ampliado a 2 años (anterior y actual).

Este es el primer estudio de validación clínica de agrupadores de morbilidad poblacional; a pesar de que agrupadores como los CRG o los ACG ya llevan muchos años en el mercado, solo se ha encontrado una referencia a un estudio de validación clínica, en curso, de los ACG por el Servicio Andaluz Salud²⁵. De los CRG existe alguna publicación en que se valida su aplicación en los sistemas de financiación² u otras validaciones de los agrupadores centradas en la capacidad explicativa sobre el consumo de recursos²⁶⁻²⁹ o la capacidad predictiva diagnóstica o pronóstica³⁰. Existe una validación clínica por médicos de AP de un listado de pacientes complejos proporcionado por un algoritmo creado con datos administrativos^{31,32}.

Un estudio llevado a cabo en el ámbito de AP por la Comunidad de Madrid con pacientes crónicos consideró que a los médicos de familia, en la mayoría de los casos, el GMA les resultaba una herramienta útil como prueba de cribado de preselección en niveles de riesgo^{1,23}.

En conclusión, los GMA presentan una estratificación de morbilidad comparable a la de los CRG, pero utilizando menos información (no incluyen procedimientos, prescripción farmacéutica, edad del paciente ni ámbito del diagnóstico²⁸), y en los estratos de mayor riesgo obtienen mejores resultados. Los clínicos consideran que ambos agrupadores estratifican adecuadamente a la población, aunque en la mayoría de los casos prefieren la que proporciona los GMA.

Lo conocido sobre el tema

Los agrupadores de morbilidad como los CRG y los GMA son herramientas de estratificación poblacional con gran poder explicativo.

Permiten ajustar variables clínicas e indicadores relacionados con la gestión sanitaria.

Son instrumentos útiles para identificar poblaciones de riesgo.

Qué aporta este estudio

La valoración, por parte de los médicos de AP, de la capacidad que tienen estas herramientas para estatificar a la población según carga de morbilidad.

Según el criterio clínico, los CRG y los GMA son instrumentos válidos para la estratificación de la población, pero, en general, los evaluadores prefieren los resultados obtenidos con los GMA.

Sugerencias, por parte de los evaluadores, de áreas de mejora de estas herramientas.

^j CAP Rambla, Mútua de Terrassa, Terrassa, Barcelona, España

^k CAP El Remei, EAP Vic, SLP, Vic, Barcelona, España

^l CAP Barberà del Vallès, Instituto Catalán de la Salud, Barberà del Vallès, Barcelona, España

^m CAP Olesa de Montserrat, Mútua de Terrassa, Olesa de Montserrat, Barcelona, España

ⁿ CAP Valldoreix, Mútua de Terrassa, Valldoreix, Barcelona, España

^o CAP Badia del Vallès, Instituto Catalán de la Salud, Badia del Vallès, Barcelona, España

^p CAP Terrassa Sud, Mútua de Terrassa, Terrassa, Barcelona, España

^q CAP Rubí, Mútua de Terrassa, Rubí, Barcelona, España

^r CAP Terrassa Oest, Mútua de Terrassa, Terrassa, Barcelona, España

^s CAP Centelles, EAP Osona Sud-Alt Congost, SLP, Centelles, Barcelona, España

Anexo B. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en [doi:10.1016/j.aprim.2018.09.016](https://doi.org/10.1016/j.aprim.2018.09.016).

Bibliografía

- Monterde D, Vela E, Clèries M, grupo colaborativo GMA. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Aten Primaria*. 2016;48:674–82.
- Hughes JS, Averill RF, Eisenhandler J, Goldfield NI, Muldoon J, Neff JM, et al. Clinical Risk Groups (CRGs): A classification system for risk-adjusted capitation-based payment and health care management. *Med Care*. 2004;42:81–90.
- Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *Lancet*. 2012;380:37–43.
- Nuño-Solinís R, Piñeira K, Rodríguez C, García A, Ayala M. La multimorbilidad vista desde la perspectiva de los profesionales de atención primaria. *Aten Primaria*. 2014;46 Supl 3:3–9.
- Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: Implications for understanding health and health services. *Ann Fam Med*. 2009;7:357–63.
- De Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. A critical review of available methods. *J Clin Epidemiol*. 2003;56:221–9.
- Grembowski D, Schaefer J, Johnson KE, Fischer H, Moore SL, Tai-Seale M, et al. A conceptual model of the role of complexity in the care of patients with multiple chronic conditions. *Med Care*. 2014;52 Suppl 3:S7–14.
- Loeb DF, Binswanger IA, Candrian C, Bayliss EA. Primary care physician insights into a typology of the complex patient in primary care. *Ann Fam Med*. 2015;13:451–5.
- Carrasco JL, Jover L. Métodos estadísticos para evaluar la concordancia. *Med Clin (Barc)*. 2004;122 Supl 1:28–34.
- Campo-Arias A, Herazo E. Concordancia intra- e interevaluadores. *Rev Colomb Psiquiatr*. 2010;39:424–32.
- Shoukri MM, Martin SW, Mian IUH. Maximum likelihood estimation of the kappa coefficient from models of matched binary responses. *Stat Med*. 1995;14:83–99.
- Freitas A, Santos JV, Lobo M, Santos C. Comparing comorbidity adjustment scores for predicting in-hospital mortality using

Conflicto de intereses

M. Clèries, D. Monterde y E. Vela son los desarrolladores de los grupos de morbilidad ajustados. El resto de los autores declaran no tener ningún conflicto de intereses.

Anexo A. Grupo de validación

Jordi Acezat Oliva^f, Antonio Arevalo Genicio^g, Joan Barrot de la Puente^h, Ander Burgaña Agouèsⁱ, Sílvia Cárceles Jurado^j, Joan Caselles i Rey^h, Jordi Casanovas Font^k, Lourdes Cruz Cubells^l, Rafael Cubí i Monfort^h, Silvia Falcón Vives^l, Xavier Gallego Laredo^l, Marta Garcia Pastor^m, Isabel Garcia Albas^l, M. Carmen Garcia López^l, Montserrat Gavagnac Bellsolàⁿ, Pau Hidalgo Valls^l, Milagros Iglesias Martínez^o, Jordi Jiménez Girado^p, Montserrat Llordes Llordes^p, Xavier López Lupión^q, Carles López i Arp^h, Alfonso Martín Pascual^r, Mireia Martínez Ortega^f, J. Carlos Martínez Vindel^l, José M. Mercadé Salavert^k, Àngels Moleiro Oliva^k, Cristina Molina Guasch^l, Teresa Mur Martí^p, Sílvia Narejos Pérez^s, Leonor Navarrete González^q, Daniel Panyart Sánchez^j, M. Dolores Ribero Genar^h, Anna Ripoll Ramosⁿ, Marta Serrarols Soldevila^k, Mercè Soler Guerra^l, Laura Taberner i Pinsach^h, Montserrat Teixidó Coletⁿ, Ramon Vilatimó Pujal^s, Alicia Villanueva Hernández^k, M. Carmen Yuste Marco^o

^f CAP Terrassa Nord, Consorcio Sanitario de Terrassa, Terrassa, Barcelona, España

^g CAP Turó de can Mates, Mútua de Terrassa, Sant Cugat del Vallès, Barcelona, España

^h CAP de Salt, Instituto Catalán de la Salud, Salt, Gerona, España

ⁱ CAP Sant Cugat, Mútua de Terrassa, Sant Cugat, Barcelona, España

- administrative data. En: Rocha Á, Correia A, Adeli H, Reis L, Costanzo S, editores. Recent advances in information systems and technologies. WorldCIST 2017. Advances in Intelligent Systems and Computing, volume 571. p. 324-31.
13. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. *Med Care*. 2005;43:607-15.
 14. Gil E, Ortiz H, López-Gay D, Zorrilla B, Arrieta F, Nogales P. Validez y concordancia de la historia clínica electrónica de atención primaria (AP-Madrid) en la vigilancia epidemiológica de la diabetes mellitus. Estudio PREDIMERC. *Gac Sanit*. 2014;28:393-6.
 15. Baldwin LM, Klabunde CN, Green P, Barlow W, Wright G. In search of the perfect comorbidity measure for use with administrative claims data: Does it exist? *Med Care*. 2006;44:745-53.
 16. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: An overview. *Med Care*. 2002;40:26-35.
 17. Quan H, Khan N, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Med Care*. 2002;40:675-85.
 18. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, et al. Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54:1423-8.
 19. Glass S, Gray M, Eden OB, Hann I. Scottish validation study of Cancer Registration data childhood leukaemia 1968-81-I. *Leuk Res*. 1987;11:881-5.
 20. Vidal C, Clèries M, Vela E, Salvador X, Ramis O, Puig J, et al. Estudio de precisión del registro de enfermos renales de Cataluña: calidad de las variables clínicas. *SEDYT*. 1991;XIII/1:11-8.
 21. Bellmunt-Montoya S. Validación de pruebas diagnósticas. *Angiología*. 2007;59:433-8.
 22. Brotons C, Monteserín R, Martínez M, Sellarès J, Baulies A, Fornasini M. Evaluación de la efectividad de un instrumento para identificar problemas sociales y sanitarios en la población anciana adscrita a un centro de atención primaria. *Aten Primaria*. 2005;36:317-27.
 23. González González AI, Miquel Gómez AM, Rodríguez Morales D, Hernández Pascual M, Sánchez Perruca L, Mediavilla Herrera I. Concordancia y utilidad de un sistema de estratificación para la toma de decisiones clínicas. *Aten Primaria*. 2017;49:240-7.
 24. Chow R, Chiu N, Bruera E, Krishnan M, Chiu L, Lam H, et al. Inter-rater reliability in performance status assessment among health care professionals: A systematic review. *Ann Palliat Med*. 2016;5:83-92.
 25. Servicio Andaluz de Salud. Consejería de Salud. Estrategia para la renovación de la Atención Primaria. Sistema ACG: Clasificación de pacientes en atención primaria [consultado 25 Mar 2018]. Disponible en: http://www.juntadeandalucia.es/servicioandaluzdesalud/principal/documentosAcc.asp?pagina=gr_renovacionap_acg
 26. Inoriza JM, Coderch J, Carreras M, Vall-Ilosera L, García-Goñi M, Lisbona JM, et al. La medida de morbilidad atendida en una organización sanitaria integrada. *Gac Sanit*. 2009;23:29-37.
 27. Sicras-Mainar A, Navarro-Artieda R, Grupo ACG-BSA. Validación retrospectiva del Johns-Hopkins ACG Case-Mix System en población española. *Gac Sanit*. 2009;23:228-31.
 28. Monderde D, Vela E, Clèries M. Validación de los grupos de morbilidad ajustados en el ámbito de la atención primaria: comparativa con los clinical risk groups. *Aten Primaria*. 2019. <https://doi.org/10.1016/j.aprim.2017.09.012>.
 29. Estupiñán-Ramírez M, Tristanco-Ajamil R, Company-Sancho MC, Sánchez-Janáriz H. Comparación de modelos predictivos para la selección de pacientes de alta complejidad. *Gac Sanit*. 2019;33:60-5. <http://dx.doi.org/10.1016/j.gaceta.2017.06.003>.
 30. Debray TPA, Vergouwe Y, Koffigberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68:279-89.
 31. Bellini I, Bartetta V, Profili F, Bussotti A, Severi I, Isoldi M, et al. Identifying high-cost, high-risk patients using administrative databases in Tuscany, Italy. *Biomed Res Int*. 2017;2017:9569348. Disponible en: <https://doi.org/10.1155/2017/9569348>.
 32. Bohórquez P, Nieto MD, Pascual B, García-Lozano MJ, Ortiz MA, Bernabéu M. Validación de un modelo pronóstico para pacientes pluripatológicos en atención primaria: estudio PROFUND en atención primaria. *Aten Primaria*. 2014;46 Supl 3:41-8.