

## Los mejores hospitales. Entre la necesidad de información comparativa y la confusión

**Salvador Peiró**

*Escuela Valenciana de Estudios para la Salud (EVES)*

Correspondencia

Salvador Peiró

EVES

Juan de Garay 21

46017 Valencia

E-mail: peiro\_bor@gva.es

Tel.: 963 869 363

Fax: 963 869 370

### Resumen

**Antecedentes:** La reciente publicación de un informe sobre los mejores hospitales españoles, y su importante repercusión mediática, ha renovado el debate sobre la utilidad de los indicadores de eficiencia y calidad hospitalaria basados en el Conjunto Mínimo de Datos Básicos (CMBD) y sobre las clasificaciones de mejores y peores hospitales a partir de estos indicadores.

**Material y método:** Revisión narrativa de los marcos conceptuales para el ajuste por gravedad y comparación de hospitales a partir de bases de datos clínico-administrativas, de las limitaciones de los sistemas y de las fuentes de datos, concretando ocasionalmente en la clasificación *top-20*.

**Resultados:** La comparación de resultados hospitalarios requiere un marco conceptual sólido que tenga en cuenta la elección de resultados claramente indicativos de mejor o peor calidad, la elección de variables de ajuste adecuadas, una cuidada metodología estadística y una presentación de los resultados informativa, así como tener en cuenta las limitaciones de los sistemas de información tipo CMBD. Los *top 20* incumplen algunos de estos supuestos haciendo dudosa la clasificación de hospitales elaborada.

**Conclusiones:** La comparación de resultados hospitalarios es una necesidad para la mejora de la calidad hospitalaria; para cumplir este objetivo deben mejorarse los métodos actualmente utilizados.

**Palabras clave:** Indicadores de calidad. Ajuste de riesgos. Índices de gravedad. Calidad de datos.

### Summary

**Background:** The recent publication of a report on the best Spanish hospitals, and their important repercussion in the media, has updated the debate about the usefulness of indicators of efficiency and hospital quality obtained from the Minimum Basic Data Set (MBDS). These indicators might let us classify better and worse hospitals in our setting.

**Material and methods:** A narrative description of the framework for risk adjustment and hospital comparison from administrative databases has been included. System and sources of data limitations have been considered, summing up occasionally in the classification *top-20*.

**Results:** The comparison of hospital outcomes needs a solid conceptual framework that keeps in mind the selection of outcomes that clearly indicate of better or worse quality. The election of adequate adjustment variables, statistical methodology and an informative presentation of the results, keeping in mind the limitations of the information systems as CMBD have been analyzed. The *top 20* has some limitations that make doubtful the hospitals' classification.

**Conclusions:** The comparison of hospital outcomes is a useful tool for the improvement of hospital quality, but to complete this objective the methodology ought to be improved.

**Key words:** Quality indicators. Risk assessment. Severity of illness index. Data quality.

### Nota: Conflicto de intereses

La Escuela Valenciana de Estudios para la Salud (EVES) es un centro dependiente de la Conselleria de Sanitat de la Generalitat Valenciana, institución que fue criticada en los medios de comunicación locales por que ninguno de sus hospitales obtuvo posiciones *top-20* en el informe de lasist-HCIA. El autor ha manifestado no tener ninguna responsabilidad sobre estos centros, ni sobre otras áreas de política o gestión sanitaria, salvo las relativas a docencia e investigación en servicios de salud en la EVES y no tener relaciones comerciales ni de asesoría con lasist-HCIA, ni con otras consultorías dedicadas a temas similares. El autor es también patrono de la Fundación Instituto de Investigación en Servicios de Salud y de la Fundación Avedis Donabedian; esta última ha colaborado con lasist-HCIA en el desarrollo de programas de monitorización de calidad hospitalaria en los que el autor no ha participado. La responsabilidad sobre los contenidos de este artículo corresponde exclusivamente al autor; ni la EVES, ni las otras instituciones citadas comparten necesariamente las opiniones expresadas en el mismo.

## Introducción

La reciente publicación de un informe sobre los mejores hospitales españoles, los *top-20* en la terminología empleada por sus autores<sup>1</sup>, y su importante repercusión mediática, ha renovado el debate sobre la utilidad de los indicadores de eficiencia y calidad hospitalaria basados en el Conjunto Mínimo de Datos Básicos (CMBD) y, sobre todo, de las clasificaciones de mejores y peores hospitales a partir de estos indicadores. Dado que nadie discute la necesidad de evaluar la eficiencia y la calidad de la atención hospitalaria, ni de disponer de información comparativa sobre los resultados de los centros, el debate real -más allá de la retórica- no se plantea en términos de defender la evaluación por indicadores más o menos objetivos frente a la ausencia de información, sino en la validez de indicadores concretos para identificar lo que pretenden identificar (los mejores hospitales), y en su utilidad en nuestro contexto.

Los centros sanitarios pueden aprender, y mucho, comparándose entre sí e intentando conocer las causas de las variaciones detectadas, tanto si se trata de variaciones en resultados clínicos (mortalidad, complicaciones, reingresos, etc.), como en utilización de recursos (duración de la estancia, etc.) o en las tasas poblacionales de ingresos o intervenciones<sup>2</sup>. La idea inicial que subyace en la comparación de estos indicadores es considerar “peores” a los hospitales con mayor duración de la estancia, mayor mortalidad, etc., pero como sabemos que estos peores resultados pueden deberse al tipo y características

de los pacientes asistidos en cada centro, y también sabemos que los hospitales tratan pacientes muy diferentes, conceptualmente consideramos legítima la variabilidad que deriva de las diferencias en los pacientes atendidos y atribuimos exclusivamente la parte de variabilidad no explicada por las características de los pacientes a diferencias en calidad de los proveedores sanitarios. La estrategia de comparación requiere, por tanto, que se ponderen los resultados de cada hospital por las características de los pacientes que ha atendido (estrategias de ajuste de riesgos).

En este trabajo se describe un marco conceptual para la evaluación comparativa de los resultados de los centros hospitalarios, incluyendo el problema de las diferencias en el tipo de pacientes atendidos y los métodos de ajuste para controlar estas diferencias, y las limitaciones genéricas para abordar estas comparaciones a partir del CMBD, concretando ocasionalmente en algunos aspectos de los *top-20*. Finalmente, se discute la utilidad de las medidas de resultados y algunas perspectivas para desarrollar métodos de comparación entre hospitales.

## Marco conceptual para la evaluación comparativa de resultados hospitalarios

### Resultados, gravedad y sistemas de ajuste

Los resultados (*outcomes*) son aquellos cambios, favorables o adversos, en el estado de salud actual o potencial de

Tabla 1. Resultados de la Atención de Salud

Supervivencia o mortalidad	Mortalidad intrahospitalaria Mortalidad durante la estancia en cuidados intensivos Mortalidad a los <i>n</i> días del alta/intervención Supervivencia a los 5 años del diagnóstico
Otros sucesos adversos	Reingresos en determinados periodos o por causas específicas Infecciones nosocomiales Enfermedades iatrógenas Complicaciones post-quirúrgicas
Estabilidad de los sistemas fisiológicos	Normalización/mejora de signos clínicos: pulso, temperatura... Normalización/mejora de parámetros bioquímicos Normalización/mejora en pruebas funcionales
Estado funcional	Capacidad para el esfuerzo Capacidad para las actividades de la vida diaria Específicos: visuales, movilidad, etc.
Función psico-social	Puntuaciones en escalas de función cognitiva, mental, ansiedad, depresión, etc.
Calidad de vida relacionada con la salud	Medición de utilidades Índices de calidad de vida Perfiles de salud
Satisfacción	Puntuación en encuestas de satisfacción
Consumo de recursos y costes	Estancia media por proceso Costes por proceso Intensidad terapéutica Tiempo de enfermería por estancia Costes por nivel de dependencia

personas, grupos o comunidades atribuibles a la atención sanitaria recibida<sup>3</sup>. En la Tabla 1 se resumen los resultados más empleados en atención hospitalaria (mortalidad, complicaciones, reingresos, duración de la estancia, etc.) con algunos ejemplos de indicadores<sup>4</sup>. La comparación de resultados es la forma usual de evaluar la eficacia de los medicamentos, pruebas diagnósticas y tecnologías, típicamente mediante ensayos clínicos con asignación aleatoria de pacientes a dos o más grupos; la idea básica que subyace en la comparación es que las diferencias -estadísticamente significativas- en resultados para los pacientes habrán sido causadas por diferencias en la eficacia de los tratamientos o las tecnologías. La asignación aleatoria de los pacientes es esencial en este tipo de estudios, ya que garantiza que quienes recibieron un determinado tratamiento no eran distintos de los que recibieron el tratamiento alternativo y, excluidas las diferencias entre pacientes como causa de las diferencias en resultados, permite atribuir los mejores resultados a la superioridad de un tratamiento sobre la alternativa (Figura 1, comparación de tecnologías).

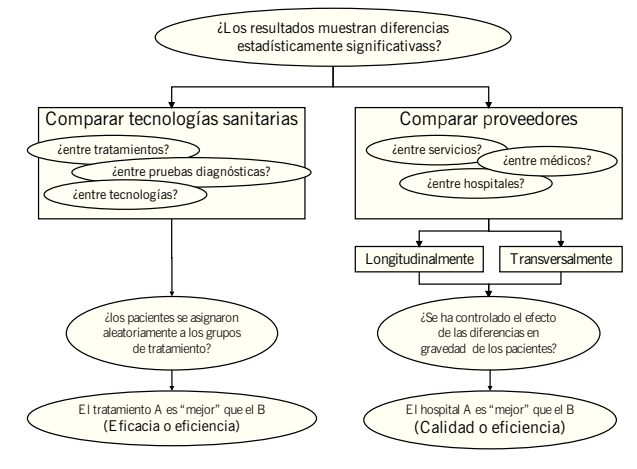
La comparación de resultados entre centros sanitarios fue un área de investigación académica durante muchos años, hasta que la introducción de los Grupos de Diagnósticos Relacionados<sup>5</sup> y, sobre todo, la publicación en 1987 -y hasta 1993- de las tasas mortalidad hospitalaria por la *Health Care Financing Administration* (HCFA)<sup>6</sup> despertó una atención sin precedentes hacia los datos de mortalidad y otros resultados médicos. El problema fundamental al comparar los resultados de diversos hospitales estriba en que, como es obvio y a diferencia de los ensayos clínicos, los pacientes no han sido asignados aleatoriamente al tratamiento por uno u otro hospital sino que, en muchos casos, han sido derivados hacia uno u otro hospital en función de su condición clínica (por ejemplo, un paciente con sospecha de hemorragia cerebral será probablemente derivado a un hospital terciario con servicio de neurocirugía). En esta situación no puede presumirse que las diferencias en mortalidad, complicaciones o duración de la estancia se deban a la

calidad de los hospitales, ya que estarán muy influidas por las características de los pacientes que han atendido (en el ejemplo, es esperable una mayor mortalidad en el hospital que recibe los casos con sospecha de hemorragia cerebral que en aquellos que los derivan, ya que el primero ingresa pacientes con elevada probabilidad de muerte y el segundo sólo se queda los casos de menor riesgo). Las características de los pacientes que definen su riesgo de peores resultados -al ingreso- se agrupan bajo la denominación de gravedad (*severity*), concepto que, en esta acepción, no debería ser interpretado sólo como riesgo de muerte, sino como probabilidad previa (derivada de las características del paciente y no de la atención recibida) de obtener el resultado que se está midiendo, que puede ser mortalidad, duración de la estancia o cualquier suceso adverso<sup>7</sup>. Este marco puede representarse mediante un esquema (Figura 1, comparación de proveedores) en el que el riesgo derivado de las características previas del paciente (gravedad), más la calidad del proveedor, conducirán a diferentes resultados y, en sentido inverso, las diferencias en resultados permitirán inferir hacia la calidad del hospital si se ha controlado el efecto de la gravedad.

Las características de los pacientes que configuran la gravedad incluyen factores como<sup>7,8</sup> la *edad*; el *sexo*; el *diagnóstico principal* y su *gravedad* (bajo el mismo diagnóstico principal de, por ejemplo, un cáncer, el pronóstico y las necesidades de recursos pueden variar extraordinariamente en función del estadiaje TNM. Hay que señalar que puede ser muy complejo dilucidar -sobre todo a partir de bases de datos con información retrospectiva- si la gravedad del diagnóstico se debe a la situación del paciente o a un manejo inapropiado de su enfermedad: ¿un absceso complicando una apendicitis estaba ya cuando el paciente llegó a urgencias o tuvo relación con la demora en el diagnóstico y tratamiento?); la *comorbilidad* (enfermedades concomitantes no relacionadas con el diagnóstico principal), incluyendo su gravedad y extensión; las *complicaciones* (patologías aparecidas durante o tras el proceso de atención, que no estaban presentes al ingreso) que pueden utilizarse como resultado en sí mismas o como factor de gravedad respecto a otros resultados como la muerte o el reingreso; la *estabilidad clínica*, que refleja el estado fisiopatológico de los sistemas corporales del paciente mediante el examen de sus signos vitales, parámetros analíticos y gases, nivel de conciencia y otras variables, y que posiblemente es el factor más crítico respecto al riesgo de muerte a corto plazo; el *estado funcional*; el *estado psicológico, cognitivo y psico-social*; las *características socio-culturales* y otras que se les asocian (nivel socio-económico, de estudios, grupo étnico, riesgos laborales, alcoholismo, tabaquismo y otras toxicomanías, hábitos alimentarios y estilos de vida, etc.); así como las *preferencias y actitudes del paciente* (rechazo de cuidados por su agresividad o por motivos religiosos, incumplimiento terapéutico, etc).

La forma de abordar el problema de si las diferencias en resultados se deben a la diferente gravedad de los pacientes o a diferencias en calidad es la habitual en epidemiología cuando se quiere aislar el efecto de factores de confusión: ajustar (usualmente mediante la modelización multivariante) los resultados obtenidos, por aquellas características de los pacientes que definían su riesgo previo respecto al resultado que se utiliza. Si

Figura 1. Utilización de los resultados para comparar tecnologías o de proveedores sanitarios



se quieren valorar diferencias en duración de la estancia o cualquier otro resultado (muertes, reingresos, complicaciones...) entre varios hospitales, pueden construirse modelos matemáticos con las variables de interés (edad, comorbilidad, diagnóstico, estabilidad clínica, etc.) para cada uno de estos resultados; de esta forma puede obtenerse un resultado ponderado por la gravedad de los pacientes atendidos en cada centro que, hasta cierto punto, puede interpretarse como el resultado que habría tenido cada hospital si sus pacientes hubieran sido iguales en las variables incluidas (en otras palabras, el resultado teórico que habrían tenido los hospitales si hubieran atendido pacientes hipotéticos con la edad promedio de todos los hospitales, el sexo promedio, un diagnóstico promedio, etc.). Esta es la base habitual para la construcción de sistemas de clasificación de pacientes, o generalizando ya que hay sistemas que no construyen grupos de pacientes, *sistemas de ajuste de riesgos (risk adjustment systems)* que, en síntesis, no son más que métodos de cuantificar la probabilidad que tienen los pacientes de obtener un determinado resultado<sup>7</sup>.

### Construcción de sistemas de ajuste

Un sistema de ajuste de riesgos se construye a partir de: 1) el resultado de interés que se pretende ajustar por los factores de riesgo; 2) una unidad de análisis, que usualmente es el episodio de hospitalización por un determinado motivo; 3) una ventana temporal para la identificación del resultado; 4) las características del paciente que definirán su gravedad respecto al resultado que se quiere medir; y 5) un método estadístico que permita valorar si las diferencias detectadas se deben o no al azar<sup>7</sup>. Un marco conceptual sólido debe reunir estos elementos, lo que básicamente implica que:

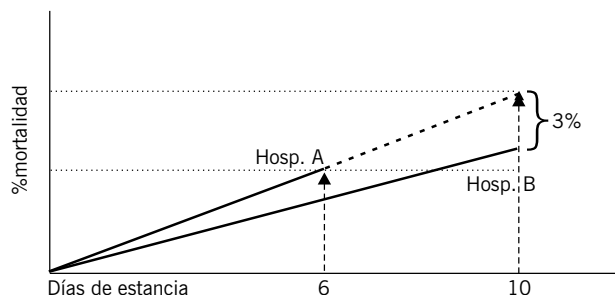
1. El resultado utilizado debe ser un indicador consistente de calidad o de eficiencia. La estancia media, por ejemplo, es un indicador aceptado de eficiencia hospitalaria, pero -y en referencia a los *top 20*- no es obvio considerar que un mayor índice de ocupación es sinónimo de mayor eficiencia o mejor gestión<sup>9</sup>. Respecto a los indicadores de calidad, tampoco es obvia la utilización de la mortalidad global como indicador de calidad; aunque la mortalidad sea un resultado adverso, en la mayor parte de las muertes hospitalarias el resultado más probable de la situación clínica del paciente era la muerte, y en estas circunstancias es confuso utilizar la mortalidad como indicador de calidad (como anécdota -extrema, pero real- que ejemplifica este problema, el hospital con peores resultados de mortalidad en la primera publicación de la HCFA -el supuestamente peor entre más de 5.000 hospitales- tenía una mortalidad del 87% cuando el modelo de ajuste le predecía un 22%; un hospital de aparente calidad pésima... de no ser porque se trataba de un hospital de terminales<sup>7</sup>).

Hay que señalar, no obstante, que algunos indicadores de mortalidad específica -como la mortalidad materna o en cirugía programada- si podrían considerarse indicadores de calidad, aunque su mínima frecuencia apenas tiene impacto en las tasas de mortalidad globales y hace complejo su análisis como tasas.

2. Los resultados deben medirse en una ventana temporal óptima para atribuirlos a problemas de calidad -una visita a urgencias a las 24 horas del alta quirúrgica sugiere más fuertemente un problema de calidad previo que la misma visita a los 6 meses- y esta ventana debe ser homogénea en todos los casos. Este último es un problema importante en los *top 20*, ya que la ventana de observación para detectar mortalidad y complicaciones (limitadas a las intra hospitalarias) varía según la duración de la estancia en cada hospital y se acaba comparando la mortalidad acumulada a los, por ejemplo, 4 días de seguimiento medio en unos centros, con la mortalidad acumulada a los, por ejemplo, 12 días en otros. En la Figura 2 se muestra un ejemplo hipotético del efecto de este sesgo. Pese a que el hospital A tiene una mayor mortalidad que el B a los 6 días de seguimiento medio (5% y 4%, respectivamente), al tener una estancia media menor, se compara la mortalidad acumulada de diferentes periodos y se ofrece un resultado erróneo. Este efecto se ha comprobado en los estudios comparativos entre mortalidad hospitalaria entre estados de EE.UU., donde las diferencias significativas en mortalidad intra hospitalaria desaparecerían si se comparaba la mortalidad a los 30 días del ingreso<sup>10</sup>. Algo parecido sucedería con las complicaciones ya que si un paciente se complica durante el ingreso el caso constará como tal, pero si es dado de alta temprana y reingresa por la misma complicación se considerará un episodio distinto.

Dado que la duración de la estancia, especialmente en los pacientes terminales, está relacionada con la disponibilidad de recursos alternativos al hospital de agudos (centros de larga estancia, unidades domiciliarias y otras) es más que probable que estos indicadores interactúen entre sí, de forma que los hospitales con estancias medias más largas tengan más mortalidad y viceversa. En

Figura 2. Efecto de las diferencias en duración de la estancia sobre la mortalidad intrahospitalaria



Hospital A: Estancia media= 6 días; Mortalidad al 6º día= 5%; Mortalidad al 10º día= 10%; Hospital B: Estancia media= 10 días; mortalidad al 6º día 4%; Mortalidad al 10º día: 8%.

el caso de los *top 20*, este problema habrá tenido el efecto de mejorar artificialmente la clasificación de mortalidad y complicaciones a los hospitales con menor duración de la estancia, de forma que a partir de un solo índice obtienen buenas puntuaciones en tres indicadores.

3. Las variables de ajuste deben ser factores de riesgo -avalados por el conocimiento científico- respecto al resultado medido. Por ejemplo, se puede utilizar el sexo para ajustar el resultado funcional de las prótesis de cadera, pero no para ajustar los resultados de las intervenciones de cataratas (porque no hay evidencias de que el sexo del paciente influya en el pronóstico en este caso). En el caso de los *top 20* este marco se quiebra notablemente al introducir como factores de ajuste de la mortalidad y las complicaciones variables del tipo "número de camas", "acreditación docente", "tamaño de la población cubierta", "comunidad autónoma". La inclusión de estas variables -posiblemente con ánimo de mejorar el rendimiento estadístico de los modelos- puede haber tenido efectos en contradicción con los resultados presentados. Por ejemplo, si -muy hipotéticamente- una Comunidad Autónoma hubiera mostrado una mortalidad agregada mayor a la de las otras Autonomías, el ajuste empleado en los *top 20* habría atribuido este exceso de mortalidad a los pacientes, y se lo habría restado a los hospitales de esa Comunidad que de esta forma saldría beneficiada frente al resto (o empeoraría en el caso de menor mortalidad); esta misma situación se habrá producido también, en alguna medida, respecto a las otras variables comentadas que no pueden considerarse como "gravedad" de los pacientes, restándose a los hospitales excesos de mortalidad "asociados" con factores como el tamaño o la docencia.

4. Los factores de ajuste deben ser características de los pacientes pero -a efectos de valorar la calidad- no deben formar parte de los procesos de atención, ni de los resultados finales. Por ejemplo, si se ajusta la mortalidad en el infarto por el porcentaje de pacientes trombolizados y el tiempo hasta la trombolisis, las diferencias en calidad desaparecerían al ser tratadas como factores de gravedad; del mismo modo, algunos diagnósticos (por ejemplo: parada cardio-respiratoria, fallo multi-orgánico, etc.) tienen una correlación tan elevada con la mortalidad que, prácticamente, significan lo mismo. Si se utiliza este tipo de diagnósticos como factores de gravedad para ajustar la mortalidad (el manual de los *top-20* lo sugiere, pero la escasa información metodológica contenida no permite afirmarlo) se habrán hecho desaparecer los factores de calidad de la atención que llevaron a, por ejemplo, la parada cardíaca. Nótese, que si el paciente ingresó con una parada cardíaca, este diagnóstico sí debería ser utilizado para ajustar la mortalidad (ya que corresponde a gravedad del paciente al ingreso), pero si ingresó para cirugía electiva y tuvo una parada cardíaca, no debería utilizarse en el ajuste, ya que la parada puede haber sido el resultado de problemas de calidad de la atención que es justamente lo que se pretende dejar como variación residual para atribuir-

selo al hospital (y no al paciente). Estos aspectos son muy complejos de diferenciar a partir del CMBD, por lo que las estrategias clásicas de ajuste optan por excluir este tipo de diagnósticos (salvo que sean el principal, por definición, el que motivó el ingreso), aspecto no obvio en el *top 20*.

Los parámetros y el marco predefinidos definirán los requerimientos de información, el momento de la toma de datos y otros aspectos del sistema de ajuste. En términos generales se dispone de 2 tipos de sistemas (Tabla 2): los derivados a partir de datos del paciente o la historia clínica -los prototipos serían el APACHE II o el *Disease Staging*- obtenidos de forma prospectiva, usualmente destinados a ajustar resultados clínicos y con grandes requerimientos de información; y los derivados a partir de bases de datos tipo CMBD, usualmente destinados a ajustar resultados de consumo de recursos, cuyo prototipo serían los GDR y que requieren relativamente poca información, que se obtiene retrospectivamente (al alta). Estas diferencias tienen implicaciones en coste y factibilidad de los sistemas, pero también en su capacidad de ajuste y su credibilidad clínica, ya que las bases de datos clínico-administrativas (*administrative databases*), como el CMBD, tienen importantes limitaciones que van desde la ausencia de variables relevantes, la toma de datos retrospectiva, los problemas de calidad de la información y la variabilidad en la cumplimentación y en la codificación diagnóstica entre centros, las dificultades para analizar pacientes a partir de bases de episodios de hospitalización y las dificultades legales para la combinación de registros (*linkage*).

### Limitaciones de las comparaciones entre hospitales basadas en el CMBD

Pese a las limitaciones señaladas, el CMBD tiene la ventaja primordial de existir -es prácticamente obligatorio, aunque en los hospitales privados el desarrollo es todavía muy bajo- e incluir un gran número de pacientes. Por ello, en nuestro país,

Tabla 2. Características de los sistemas de ajuste de riesgos

Fuente de datos	Paciente e historia clínica	Bases de datos clínico-administr.
Resultado a ajustar	Clínicos	Consumo recursos
Requerimientos de información	++++	+
Coste de funcionamiento	++++	+
Capacidad ajuste	+++	+
Credibilidad clínica	++	+
Desarrollo	Juicio Clínico	Empírica
Importancia del diagnóstico	++	++
Importancia del procedimiento	+	++++
Importancia variables fisiológicas	+++	-
Toma de datos	Variable	Al alta

CMBD: Conjunto Mínimo de Datos Básicos.

el CMBD es la fuente de información esencial -casi única- para la construcción de sistemas de ajuste de riesgos que permitan la comparación de hospitales. En este apartado se describen las principales limitaciones de las comparaciones entre hospitales desarrolladas a partir de datos del CMBD, la fuente esencial empleada en los *top-20*, utilizando un marco de clasificación según estas limitaciones deriven de: 1. los sistemas de codificación diagnóstica, 2. la estructura y fuentes de datos del CMBD, 3. la variabilidad de las prácticas hospitalarias de cumplimentación de los datos; 4. los modelos de ajuste de riesgos y, 5. las formas de tratamiento y presentación de la información final.

### **Limitaciones de la Clasificación Internacional de Enfermedades**

El CMBD utiliza para la codificación de diagnósticos y procedimientos la Clasificación Internacional de Enfermedades 9ª revisión Modificación Clínica (CIE9MC), una adaptación para uso clínico de la Clasificación Internacional de Enfermedades (CIE), originalmente desarrollada para clasificar causas de muerte. La principal crítica se centra en la falta de definiciones clínicas operativas, que se traduce en la inclusión bajo los mismos códigos de pacientes muy diferentes<sup>11,12</sup>. En la tabla 3 se muestra un ejemplo del impacto de este problema: el hospital V diagnosticó el 2% de sus partos como sufrimiento fetal, frente al 16% del hospital IV, pero el primero practicó cesárea en el 97% de estos partos por sólo el 7% de cesáreas en el segundo, sugiriendo que estos hospitales etiquetan como sufrimiento fetal pacientes muy diferentes. Nótese que este problema no es tan importante para casos de menor dificultad diagnóstica como el parto de nalgas<sup>13</sup>. A efectos del ajuste de riesgos es un problema crítico ya que los resultados del segundo hospital serán valorados como si sus casos tuvieran la misma gravedad real que los del segundo, y se le justificarán tasas más altas de cesáreas o mortalidad (ya que tenía más casos de supuesto sufrimiento fetal) que al otro. La CIE9MC también ha sido criticada<sup>7,11,12</sup> por su orientación mono-axial -derivada de su diseño original orientado a determinar la causa principal de muerte- que no recoge muchas interacciones entre diagnósticos, ni determinadas características de los pacientes de importancia pronóstica reconocida. Incluso en algunas enfermedades en que la clasificación ha realizado desarrollos para incorporar complica-

ciones y afectaciones multiorgánicas, el estado real del paciente puede ser difícil de discernir.

Los criterios esenciales para realizar un diagnóstico son la etiología, la localización y las manifestaciones fisiopatológicas (por ejemplo, neumonía de lóbulo inferior izquierdo, por *s. pneumoniae* con insuficiencia respiratoria). Sin embargo, en la práctica clínica no siempre es posible (o deseable) llegar más allá de un diagnóstico sindrómico y la adaptación a esta realidad clínica ha resultado en el desarrollo de clasificaciones diagnósticas no mutuamente excluyentes, en las que el mismo paciente puede ser clasificado de forma diversa en función de la información disponible, aspecto que introduce nuevas fuentes de variabilidad en la asignación diagnóstica en función de la disponibilidad de medios diagnósticos y su uso más o menos intensivo. En la Figura 3 se muestra un ejemplo de posibilidades de uso de diferentes etiquetas diagnósticas en los mismos pacientes, que se trasladarían a diferentes códigos CIE9MC y a un peso relativo de reembolso en el sistema GDR muy diferente. Relacionado con los anteriores, se hallan los problemas vinculados a las guías de codificación, manuales que intentan trasladar el lenguaje clínico usual -vago y sujeto a incertidumbre- a códigos CIE9MC, pero que aumentan la fiabilidad del CMBD a expensas de sesgos de mala clasificación (p. ej.: un ingreso para biopsia hepática que es dado de alta con el diagnóstico de "sospecha de cáncer de hígado", será codificado como hepatocarcinoma, incluso si el diagnóstico anatómico-patológico -usualmente disponible varios días tras el alta- no se confirma).

En conjunto, estos problemas, derivados de la propia CIE9MC y que afectan tanto al *top-20* como a cualquier otro sistema basado en códigos CIE9MC, se van a traducir en amplias posibilidades de clasificar tanto pacientes similares bajo diferentes epígrafes, como pacientes muy diferentes bajo un mismo epígrafe. Se trata de un aspecto difícilmente salvable que favorece otros problemas como la búsqueda de las asignaciones de códigos más rentables. Las implicaciones sobre la comparación de centros son obvias.

### **Limitaciones derivadas de la estructura del CMBD**

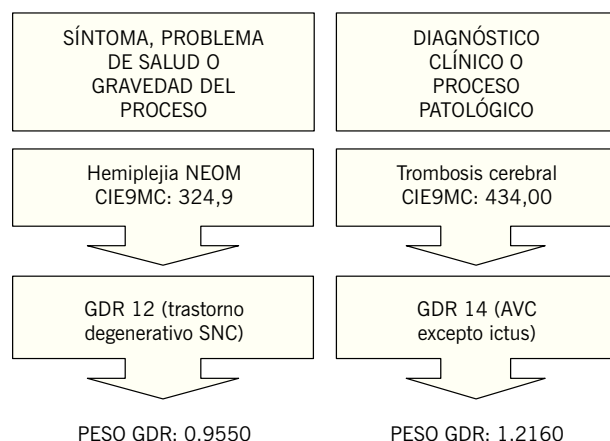
Respecto a la estructura de los CMBD, un primer problema a tener en cuenta es la limitación del número de diagnósti-

Tabla 3. **Porcentaje de partos con determinados riesgos y de cesáreas según tipo de riesgos en hospitales seleccionados**

Hospital	Nalgas		Distocia		Sufrimiento Fetal	
	% Nalgas	% Cesárea	% Distocia	% Cesárea	% Sufrimiento fetal	% Cesárea
I	1,62	61,54	6,67	91,59	2,28	86,30
II	4,53	65,91	7,78	97,53	14,98	27,86
III	1,77	90,00	18,76	47,17	10,97	95,16
IV	2,37	70,31	7,64	97,09	16,05	7,39
V	0,89	40,00	7,45	96,83	2,19	97,30
Media	2,96	70,28	8,11	85,86	8,38	35,19

Basado en: Librero J, *et al.*<sup>13</sup>

Figura 3. Efecto de las diferencias en especificidad diagnósticas sobre el reembolso (ejemplo)



Paciente de 86 años que ingresa por urgencias con hemiparesia en mano y pierna derecha y afasia. Durante el ingreso se muestra clínicamente estable y al 4º día es dado de alta para seguimiento en consulta externa.

NEOM: no especificado de otro modo; CIE9MC: Clasificación internacional de enfermedades 9ª revisión modificación clínica. GDR: Grupos de diagnósticos relacionados; AVC: accidente vascular-cerebral; SNC: sistema nervioso central.

cos secundarios (9 a 12 usualmente en los CMBD españoles), tal vez suficiente para los procesos agudos pero incapaz de recoger la complejidad de los pacientes con comorbilidad múltiple o con complicaciones, aspectos de reconocida importancia pronóstica y respecto a los costes. Se ha señalado que los cuadros crónicos son los más obviados por los clínicos (o los codificadores), dando lugar a resultados paradójicos en los que la presencia de comorbilidad crónica -sólo recogida cuando el

paciente no tiene comorbilidad aguda- se comporta como factor protector de mortalidad intrahospitalaria<sup>9,14-16</sup>. En la Tabla 4 se muestran ejemplos de este sesgo en CMBD españoles, donde la presencia de un diagnóstico secundario de, por ejemplo, hipertensión arterial, se asociaba a una reducción del 50% en el riesgo de muerte en el infarto agudo de miocardio. Hay que señalar, que este problema se ha registrado incluso en los CMBD de California que recogen hasta 25 diagnósticos secundarios, indicando la existencia de un sesgo de información no atribuible exclusivamente a la falta de campos para la anotación de diagnósticos secundarios en el CMBD<sup>14</sup>.

Un segundo problema se refiere a la ausencia de variables de importancia pronóstica reconocida. Especial relevancia tienen en este sentido la ausencia de variables de estabilidad fisiopatológica (pulsaciones por minuto, frecuencia respiratoria, presión arterial, pH, pO<sub>2</sub> y pCO<sub>2</sub>, hematocrito y demás), claves en los sistemas de ajuste respecto a la mortalidad inmediata utilizados en las unidades de cuidados intensivos (APACHE II y III, MPM, etc.). También cabe destacar la ausencia de otras variables de interés que afectan al resultado o de algunos resultados que pueden ser los relevantes en algunas patologías; nótese que de las variables de gravedad enumeradas previamente, el CMBD sólo trabaja con la edad, sexo, diagnóstico principal y secundarios (mezcla de comorbilidad y complicaciones) y una variable de proceso -procedimientos- que es muy difícil de interpretar en términos de calidad.

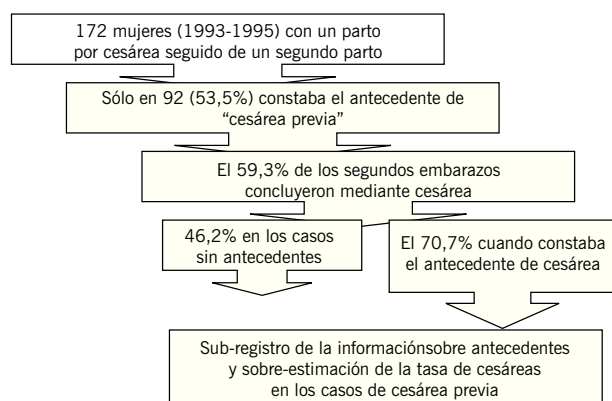
Un último problema se refiere a la naturaleza retrospectiva -al alta- de la toma de datos, que no permite distinguir si se trata de cuadros tratados durante la hospitalización o simplemente son procesos coexistentes (la valoración de la calidad varía si un paciente que ingresó para intervención tenía una infección al ingreso o la desarrolló en el hospital)<sup>17-20</sup>. Aunque se han utilizado interesantes fórmulas para enfrentar este problema con técnicas cualitativas y algoritmos<sup>17</sup>, todavía se está lejos de una solución óptima<sup>20-22</sup>. El sesgo de información retrospectiva también se traduce en la anotación o no de determinados factores en función de los propios resultados. Por ejemplo, los pacientes que tienen alguna complicación grave tienen

Tabla 4. Asociación entre mortalidad hospitalaria y registro de diagnósticos secundarios (Odds ratios con p<0,05, ajustadas por edad, sexo, días de estancia y número de diagnósticos registrados)

Título	AVC	NEM	IAM	ICC
Diabetes mellitus no complicada ni insulindep.	0,53	0,49	ns	ns
Hipertensión arterial NE	0,49	0,54	0,52	0,67
Enfermedad cardíaca hipertensiva NE, sin ICC	ns	ns	0,82	0,44
Infarto antiguo de miocardio	ns	ns	0,35	ns
Otra angina de pecho y angina de pecho NE	ns	ns	0,44	ns
Enfermedad cardíaca isquémica crónica NE	ns	ns	ns	0,42
Alteraciones de la válvula mitral	ns	ns	ns	0,57
Alteraciones de la válvula aórtica	0,47	ns	ns	ns
Fibrilación auricular	0,68	ns	0,36	0,53
Otras obstrucciones crónicas respiratorias	ns	0,56	0,38	ns
Infección de vías urinarias sitio NE	0,44	0,08	ns	ns

AVC: Accidente vascular cerebral; NEM: neumonía; IAM: infarto agudo de miocardio; ICC: insuficiencia cardíaca congestiva. NE: no especificado; ns: no significativo. Basado en: Librero J, et al.<sup>15</sup>

Figura 4. **Sesgo de información en bases de datos administrativas (ejemplo)**



Basado en Librero J, et al.<sup>13</sup>

una mayor probabilidad de tener registrada información que justifique esta complicación que aquellos pacientes con el mismo cuadro que no desarrollaron la complicación. En la Figura 4 se muestra un ejemplo de cómo a las pacientes que se les practicó una cesárea tras haber tenido una cesárea previa, tenían en mayor proporción registrado el antecedente que aquellas a las que en las mismas circunstancias no se les practicó cesárea<sup>13</sup>.

### Prácticas hospitalarias y manipulaciones interesadas

Las limitaciones anteriores permiten que los hospitales desarrollen sus bases de datos dentro de un amplio margen de variabilidad e, incluso, la manipulación de los sistemas de información con las finalidades de mejorar la posición relativa del hospital o mejorar el reembolso<sup>23</sup>. La vaguedad en la identificación del diagnóstico principal, el diseño de la CIE9MC que no considera interrelaciones entre códigos y las posibilidades de seleccionar secundarios, estimulan la búsqueda de las asignaciones de códigos más rentables para el hospital. Simborg definió como *DRG-creep* la práctica desarrollada por muchos hospitales de seleccionar los diagnósticos y sus combinaciones de modo que se maximizarán el reembolso económico<sup>24</sup>, y también se ha definido como *upcoding* la práctica, evidenciada en diversos estudios, de incluir diagnósticos más o menos dudosos que permiten el cambio de casos de un GDR sin complicaciones a su pareja complicada con mayor peso de reembolso (esta práctica parece haber disminuido tras las inspecciones realizadas por el *General Accountig* -acompañadas de fuertes sanciones a los hospitales- y la difusión de guías de codificación para el reembolso con criterios estrictos<sup>25</sup>), pero se trata de un problema poco explorado en España donde las inspecciones sobre la codificación son mínimas y los productos derivados del CMBD apenas se utilizan para el reembolso (salvo en la Comunidad, que incluyó 13 hospitales en el *top-20*). En todo caso, si es conocida la amplia variabilidad en la codificación hospitalaria en España<sup>26</sup> y la presencia de cambios de asignación del GDR entre el 20% y 30% de los casos (entre 1 de cada 5 y 1 de

cada 3 altas) cuando se realizan revisiones de la codificación diagnóstica<sup>27,28</sup>.

### Limitaciones derivadas de los sistemas de ajuste empleados

La validez final y la utilidad práctica de los ajustes viene dada por su capacidad para capturar las diferencias entre pacientes, problema que es abordado usualmente comparando las predicciones realizadas por los modelos ajustados con los resultados reales. La modelización multivariante proporciona el marco para el traslado de puntuaciones de riesgo, categorías de riesgo o variables específicas a resultados esperados. Las variables independientes pueden ser cualquiera de los factores de riesgo comentados anteriormente, y la dependiente es el resultado a analizar que, básicamente, puede ser una variable continua (días de estancia, costes) o dicotómica (presencia o no de muerte, complicación, reingreso). Estadísticamente, suelen emplearse modelos pertenecientes a la familia de los Modelos Lineales Generalizados, típicamente modelos de regresión lineal múltiple cuando la variable resultado es continua, y de regresión logística cuando la variable resultado es dicotómica.

El coeficiente de determinación  $R^2$  es la medida de rendimiento de los modelos multivariantes mas utilizada cuando la variable dependiente es continua y puede ser interpretado como la fracción de la variabilidad total de la variable dependiente que puede ser explicada por las diferencias de riesgo entre los casos incluidos en el modelo. En teoría, el modelo tendrá un mejor rendimiento cuanto más cerca esté el coeficiente  $R^2$  del valor 1, pero en la práctica el valor del  $R^2$  está en función del número de variables e interacciones introducidas en una base de datos concreta, del entorno del que proviene la muestra, especialmente del tipo de pacientes incluidos, de la dispersión de las variables, de la inclusión o no de *outliers* y de las transformaciones de variables efectuadas. Respecto al rendimiento de los GDR, el sistema empleado por los *top 20* para ajustar la duración de la estancia, se conoce<sup>29-31</sup> el mayor  $R^2$  de los casos quirúrgicos (cuyos datos ofrecen menos dispersión) y la pérdida de capacidad explicativa cuando se aplican en bases de datos diferentes a las que se utilizaron para su desarrollo. Esto sugiere que la capacidad de ajuste de los GDR varía en función del tipo de pacientes atendidos en cada centro, siendo óptima cuando el hospital tiene una importante proporción de pacientes quirúrgicos (especialmente en patologías bastante homogéneas en duración de la estancia, como las cataratas, hernias inguinales, colecistectomías, etc.), y mínima cuando el hospital tiene una elevada proporción de casos médicos. Un problema adicional es que los pesos de los GDR tienden a estar sesgados, infraestimando el riesgo para los casos de alto riesgo y sobreestimándolo en los de bajo riesgo (sesgo de compresión)<sup>32</sup> que implica que los hospitales con casos más graves verán infraestimados sus riesgos en las bases de datos integradas fundamentalmente por casos de bajo riesgo<sup>33</sup>.

La regresión logística es el método más utilizado para modelizar variables dependientes dicotómicas y parece ser el método usado para modelizar la mortalidad y las complicaciones en los *top 20*. En la evaluación del rendimiento de estos modelos se emplean dos criterios: calibración y discriminación. Un modelo se dice calibrado respecto a un conjunto de datos cuando el promedio de sus predicciones se aproxima al prome-



dio de los resultados reales (usualmente se consideran calibrados si la prueba de Hosmer-Lemeshow no es significativa, aunque esta información no se ofrece en los *top-20*). Por otro lado, un modelo discrimina si predice mayores probabilidades del suceso (ej.: muerte) a los pacientes con el suceso (los que realmente han muerto). El estadístico C -igual al área bajo la curva ROC- es el más utilizado para valorar la discriminación<sup>34,35</sup> y puede tomar valores entre 0,5 (el modelo no funciona mejor que el azar para predecir el resultado) y 1 (el modelo predice perfectamente). Como cautelas, son aplicables a estos modelos todas las planteadas anteriormente, en particular sobre las comparaciones de rendimiento cuando son obtenidas de diferentes grupos de datos, o existen diferentes criterios en definición de la población, protocolos en la depuración de datos, sobre representación de casos difíciles y, en definitiva, situaciones que influyen sustancialmente en la variabilidad de los datos que el modelo ha de explicar.

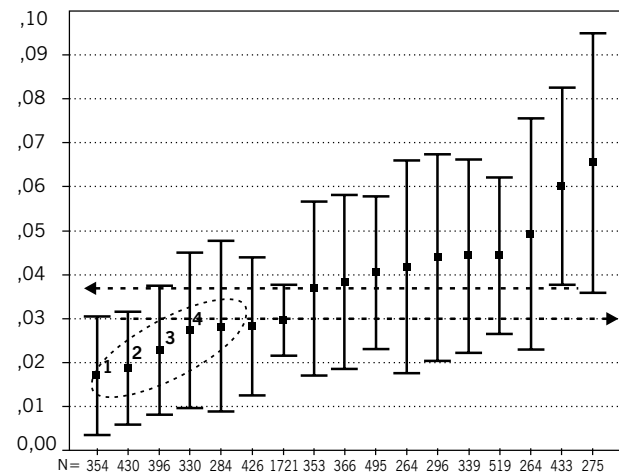
En el caso de los *top 20* no se ofrecen los resultados de la calibración, pero si el estadístico C del modelo global de mortalidad que alcanza el valor de 0,94, un valor sorprendentemente bueno para un modelo basado en CMBD (mejor que modelos con mucha más información como el APACHE II o el *Mortality Probability Model* que apenas superan el 0,80, aun sobre grupos de pacientes más homogéneos). La explicación hay que buscarla probablemente en el uso de muchas variables -saturación del modelo\* -y de variables explicativas, fundamentalmente el diagnóstico, obtenidas retrospectivamente cuando ya se conoce si el paciente ha fallecido o no y que, pese a que contribuyen a mejorar la capacidad predictiva porque correlacionan muy bien con el resultado, no permiten valorar la calidad de los hospitales.

Un problema a añadir es que el informe *top-20*, posiblemente para mantener el "secreto comercial", apenas contiene información para valorar los modelos (no se presentan los modelos con sus coeficientes, ni la significación de las variables ni otros detalles necesarios para valorar su bondad). Esto implica que los hospitales son juzgados por un sistema que ellos no pueden juzgar y, al margen de una cierta situación de indefensión, tampoco pueden conocer dónde tienen los problemas que deberían resolver para obtener mejores resultados, ni qué hacen bien los hospitales *top* para poder "copiar". Especialmente oscuro resulta también cómo se han identificado las complicaciones (un aspecto de interés ya que éste sí sería -a diferencia de la mortalidad- un indicador consistente de calidad). Esta estrategia de "caja negra"<sup>36</sup> puede tener un cierto sentido comercial, pero es poco comprensible desde el punto de vista de la calidad asistencial que, fundamentalmente, busca conocer para cambiar.

#### Limitaciones derivadas de los métodos de presentación de resultados

La presentación de los resultados, especialmente cuando se difunden públicamente es un aspecto de interés, ya que una

Figura 5. Solapamiento de los intervalos de confianza de las tasas de mortalidad en el ingresos por fractura de cadera. Hospitales de la Comunidad Valenciana 1995



Cada línea representa el porcentaje de mortalidad (desde menos del 2% a más del 6%) y su intervalo de confianza del 95%. Las líneas punteadas proyectan el límite superior del intervalo de confianza del hospital de menor mortalidad, y el límite inferior del hospital con mayor mortalidad. A pie de gráfico se muestra el número de casos en cada hospital.

presentación poco cuidadosa puede contribuir a la confusión. Existen muchas formas de realizar estas presentaciones, pero la regla esencial es hacer más hincapié en los estadísticos de dispersión que en los centrales. En la Figura 5 se muestran los valores medios y el intervalo de confianza del 95% de la tasa de mortalidad en fractura de cadera en 19 hospitales de la Comunidad Valenciana. Al solaparse los intervalos de confianza de los 17 primeros hospitales (o de los 17 últimos, si se buscan los peores hospitales) la presentación correcta debería señalar que no existen diferencias entre estos hospitales, y no que hay 4 primeros o 4 últimos (en este ejemplo sólo podría afirmarse que los dos primeros son diferentes de los dos últimos). Es posible que el sistema empleado en los *top 20* haya permitido encontrar diferencias estadísticamente significativas en duración de la estancia entre hospitales (las variables cuantitativas obtienen intervalos de confianza relativamente estrechos), pero es muy improbable que haya permitido encontrar diferencias entre hospitales -salvo quizás los extremos- en sucesos dicotómicos de baja frecuencia (como la mortalidad o las complicaciones), y ello pese al exceso de poder estadístico que deriva del uso de muestras de gran tamaño. En estas circunstancias, la ordena-

\*Si todas las combinaciones posibles de las variables predictoras se incluyen en el modelo se dice que éste está saturado, lo que se traduce en que cada caso se encuentra en una y sólo en una de las celdas definidas por las interacciones. Por ejemplo, si el modelo incluyera sólo el sexo y 5 categorías de gravedad, un modelo completamente saturado comprendería 10 posibilidades o variables (9 variables indicador y un término constante). En un modelo saturado, el valor estimado para cada celdilla es igual al valor promedio de la variable dependiente para todos los casos incluidos en ella.

ción de hospitales por sus valores medios (y la suma posterior de rangos) no parece tener sentido.

## Utilización comparativa de resultados ajustados

### ¿Son los top 20 los mejores hospitales?

Una posibilidad para estar entre los top 20 es ser un hospital realmente mejor que otros. Existen, sin embargo, otros mecanismos -más o menos legítimos- que permitirán que hospitales iguales (y tal vez peores) que otros puedan alcanzar esta cima. Entre los legítimos cabe citar el uso de diagnósticos más específicos, la anotación exhaustiva de las comorbilidades que influyen en el ingreso actual, la anotación de complicaciones (en este caso se mejoraría en los indicadores de mortalidad y estancia media, pero no en el de complicaciones) y la anotación exhaustiva de procedimientos. Este fenómeno -que implica una mayor calidad de la información pero no necesariamente del hospital, y es muy llamativo cuando algunos hospitales comarcales tienen pesos de casuística similares a los hospitales universitarios- se ha dado en mayor medida en los hospitales concertados de Cataluña, donde los GDR -y por tanto la cumplimentación del CMBD- tienen un cierto impacto en el reembolso, y esta podría ser una buena explicación para la buena posición de los hospitales de esta Comunidad en los top-20. Del mismo modo, una serie de circunstancias pueden empeorar la posición relativa de hospitales excelentes: como no disponer de centros para derivar pacientes crónicos o terminales, áreas geográficas con mayor preferencia por morir en el hospital, atender casos de mayor gravedad, tener una relativa mayor proporción de casos médicos (respecto a la cirugía electiva), atender a poblaciones marginales (los indigentes suelen fallecer en el hospital y los hospitales que atendían tradicionalmente la beneficencia pueden ser muy castigados por este factor).

También existen mecanismos ilegítimos que mejorarían la posición relativa de los hospitales: usar como diagnóstico principal aquel que maximice el reembolso (*creeping*), etiquetar diagnósticos específicos bajo criterios operativos dudosos (por ejemplo, clasificar como accidente vascular-cerebral cualquier sospecha de este cuadro, aun las no verificadas), anotar comorbilidades y antecedentes que no influyen en el proceso actual, ingresar durante 1 solo día procesos ambulatorizables, fraccionar un proceso en varios (por ejemplo, dando altas entre servicios, o en lugar de permisos de fin de semana, o fraccionar un mismo proceso en 2 ingresos), seleccionar casos más leves en el mismo GDR y algún otro. Si a esto se añade que los modelos empleados no han conceptualizado claramente los indicadores de calidad ni las variables de gravedad, el resultado final son grandes dudas sobre el lugar de los hospitales en esta clasificación y, por tanto, dudas sobre el valor de la clasificación en sí.

### Utilidad de las comparaciones ajustadas

En Estados Unidos la evidencia empírica sobre la utilidad de publicar resultados ajustados es escasa y, en general, decepcionante<sup>37</sup>. Sólo 7 sistemas han sido objeto de evaluación y, casi siempre, con estudios descriptivos y observacionales: los usuarios y las aseguradoras apenas buscan y, menos aún, utili-

zan la información de resultados; los clínicos muestran un gran escepticismo y sólo un porcentaje mínimo dice utilizarlos para derivar pacientes a hospitales (u otros especialistas) con mejores resultados; los hospitales parecen tener más en cuenta estos datos y, en algún Estado, se ha asociado la publicación con una mejora en los resultados para los pacientes, aunque mejoras similares se han conseguido en otros Estados que no utilizaban estas políticas. Adicionalmente, varios trabajos han mostrado como la posición de los hospitales en estas clasificaciones varía en función del método de ajuste utilizado<sup>38-40</sup> y basta consultar unos cuantos sistemas top -en EE.UU., existen numerosas consultorías que realizan este tipo de trabajo- para ver que los 100 mejores hospitales no coinciden entre clasificaciones.

Para bien y para mal, España no es Estados Unidos. La planificación sanitaria de base territorial implica una barrera a la libre elección de proveedores -que difícilmente pueden elegir a los supuestos mejores hospitales- y ha propiciado el trabajo hospitalario en equipos que atienden grandes volúmenes de población (probablemente estos dos aspectos son puntos fuertes del comparativamente bajo gasto sanitario y alta calidad técnica en los hospitales en España). Además, los proveedores, en su gran mayoría, son de propiedad pública y su personal goza (o sufre) de un *status quo* semi-funcionario, por lo que no pueden ser expulsados del mercado sanitario (al menos sin grandes dificultades), aunque los clientes no los elijan. Esto, unido a la no separación (más allá de los discursos) entre financiadores y proveedores hace difícil esperar que los centros -bajo la presión de la elección de los usuarios- tomen iniciativas para su corrección. En cuanto a los gestores, su consideración parece depender más de factores políticos que de su gestión (curiosamente, al menos dos gerentes de hospitales top-20 fueron cesados en las semanas posteriores a la difusión de este informe).

Sin embargo, la comparación de hospitales es útil aun en los entornos sanitarios con escasa competencia. Los sistemas diseñados para el ajuste de costes por proceso intentan agrupar a diferentes tipos de pacientes en términos de su consumo esperado de recursos con el objetivo básico de implementar incentivos para mejorar la eficiencia (en el sentido de disminuir los costes unitarios) en el tratamiento de los diferentes grupos de pacientes. Los sistemas diseñados para ajustar la efectividad de la atención hospitalaria, tienen por objetivo permitir la formulación de juicios sobre la calidad de la atención prestada. Bien contruidos, estos sistemas (y pese a las limitaciones comentadas) pueden tener interesantes usos clínicos, en gestión sanitaria y en otras áreas (Tabla 5). Una reciente reunión entre la Sociedad Española de Calidad Asistencial y el INSALUD ofrecía algunas pistas de que significa "bien contruidos"<sup>41</sup>: "...tanto los indicadores de sucesos adversos (tasas de infecciones u otras complicaciones de herida quirúrgica, tasas de reintervenciones, tasas de reingresos por determinados problemas, tasas de errores/reacciones adversas a medicamentos, etc.), como los indicadores de proceso soportados por evidencia científica (tiempo hasta la trombolisis en pacientes susceptibles, porcentaje de artroplastias cadera levantados al día siguiente de la intervención, profilaxis antibiótica en cirugía por fractura de cadera, etc.), así como indicadores de costes, listas de espera y satisfacción son susceptibles de difusión. No obstante, se recomienda evitar la difusión de resultados

Tabla 5. **Utilidad de los sistemas de ajuste de riesgos para diversos agentes del sistema sanitario**

Financiadores y compradores	<ul style="list-style-type: none"> <li>- Sistemas de pago a proveedores: incentivos a la eficiencia y calidad.</li> <li>- Evaluación comparativa de resultados de proveedores.</li> </ul>
Proveedores	<ul style="list-style-type: none"> <li>- Actividades de garantía de calidad.</li> <li>- Planificación estratégica de la producción.</li> <li>- Negociación con los compradores o terceros pagadores.</li> <li>- Estrategias de reducción de costes unitarios por proceso.</li> </ul>
Clínicos	<ul style="list-style-type: none"> <li>- Actividades de garantía de calidad.</li> <li>- Monitorización de actividad y resultados.</li> <li>- Negociación de objetivos de calidad y productividad.</li> </ul>
Pacientes	<ul style="list-style-type: none"> <li>- Facilitar la elección de centro asistencial.</li> </ul>
Salud pública	<ul style="list-style-type: none"> <li>- Vigilancia epidemiológica.</li> <li>- Seguimiento morbi-mortalidad.</li> <li>- Investigación de resultados.</li> </ul>

*fácilmente malinterpretables (por ejemplo, las tasas de mortalidad). [...] La fiabilidad y validez de la información requiere el desarrollo y mejora de los sistemas de información (la calidad de los datos) [...] También se requiere una mejora de las metodologías de ajuste de riesgos. [...] Respecto al nivel sobre el que debe ofrecerse información, existe coincidencia en que el hospital es un nivel excesivamente amplio... lo más razonable es ofrecer resultados por grupos patológicos (no necesariamente grupos de diagnósticos relacionados) o por procedimientos, a nivel de hospital".*

En este sentido, las críticas a los Top-20 en este texto no deberían verse como críticas a la idea de comparar resultados, ni siquiera a partir del CMBD (y mucho menos como una sugerencia de que otras empresas en España realizan estos estudios con metodologías mas correctas que la de lasist-HCIA que, por lo demás, emplearon una metodología aparentemente muy similar a la empleada por HCIA en EE.UU.), sino a un método concreto que clasifica a hospitales como "mejores" (lo que implica, aunque no se diga expresamente, que el resto son peores) sin una base metodológica lo suficientemente sólida, sin tener detrás una clara estrategia de mejora de calidad y sin ofrecer la suficiente información para juzgar la bondad del propio sistema de clasificación. Bien al contrario, la comparación iniciada por lasist-HCIA tiene la virtud de abrir un camino. Probablemente las claves del desarrollo útil de este tipo de comparaciones son: 1) una selección de indicadores con alto valor predictivo sobre la calidad, pero poco dramáticos y poco susceptibles de malas interpretaciones, 2) exquistez técnica, empleando los mejores ajustes posibles y centrándose más en los estadísticos de dispersión que en los centrales, de forma que los resultados de los centros -especialmente si son pequeños- se muevan en rangos antes que en valores medios, 3) acompa-

ñarlos de las explicaciones y precauciones necesarias para aquellos que tengan que usarlos, tanto si son gestores, como clínicos, pacientes o medios de comunicación, y 4) evitar transmitir al público, ni a nadie, que tales medidas son el arbitro científico de la mejor o peor calidad de la atención hospitalaria<sup>42</sup>.

## Bibliografía

1. lasist. *Top-20. Los mejores hospitales*. Barcelona: lasist, 2000.
2. Peiró S, Lorenzo S. La difusión a los ciudadanos de los resultados de la asistencia sanitaria (carta). *Rev Calidad Asistencial* 2000;15:684-5.
3. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166-206.
4. Peiró S, Librero J. Evaluación de calidad a partir del conjunto mínimo de datos básicos al alta hospitalaria. *Rev Neurol* 1999; 29:651-61.
5. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case Mix Definition by Diagnosis Related Groups. *Med Care* 1980; 18(Suppl):1-53.
6. Health Care Financing Administration. *Medicare Hospital Mortality Information: 1986 - Vol II*. Washington: Department of Health and Human Services, 1987.
7. Iezzoni LI, Ed. *Risk adjustment for measuring health care outcomes* (2 ed). Ann Arbor, Michigan: Health Administration Press, 1996.
8. Librero J, Peiró S, Ordiñana R. *Comparación de resultados, calidad y costes usando bases de datos. Quaderns de salut pública i administració de serveis de salut 13*. Valencia: Institut Valencià d'Estudis en Salut Pública, 1997.
9. Peiró S. Medidas de actividad y producto sanitario. En: del Llano Señaris J, Ortún Rubio V et al, eds. *Gestión Sanitaria. Innovaciones y desafíos*. Barcelona: Masson, 1998:197-218.
10. Jencks SF, Williams DK, Kay TL. Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. *JAMA* 1988;260:2240-6.
11. McMahon LF Jr, Smits HL. Can Medicare prospective payment survive the ICD-9-CM disease classification system? *Ann Intern Med* 1986;104:562-6.
12. Iezzoni LI. Using administrative diagnostic data to assess the quality of hospital care. Pitfalls and potential of ICD-9-CM. *Int J Technol Assess Health Care* 1990;6:272-81.
13. Librero J, Peiró S, Márquez Calderón S. Inter-hospital variations in caesarean sections. A risk adjusted comparison in the Valencia public hospitals. *J Epidemiol Community Health* 2000; 54:631-6.
14. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality?. *JAMA* 1992;267:2197-203.
15. Librero J, Peiró S. ¿Previenen las enfermedades crónicas la mortalidad intrahospitalaria? Paradojas y sesgos en la información sobre morbilidad hospitalaria. *Gac Sanit* 1998;12:195-202.
16. Librero J, Peiró S. ¿Paradojas de la información sobre morbilidad hospitalaria?. *Gac Sanit* 1999 (carta);13:75-7.

17. Iezzoni LI, Daley J, Foley SM, Heeren T, Hughes JS, Fisher ES, *et al.* *Screening Quality of Care Using Administrative Data. Final Report AHCPR Grant RO1-HSO6512.* Boston, MA: Boston's Beth Israel Hospital, 1992.
18. Iezzoni LI, Daley J, Heeren T, Foley SM, Fisher ES, Duncan C, *et al.* Identifying complications of care using administrative data. *Med Care* 1994;32:700-15.
19. Iezzoni LI, Daley J, Heeren T, Foley SM, Hughes JS, Fisher ES, *et al.* Using administrative data to screen hospitals for high complication rates. *Inquiry* 1994;31:40-55.
20. Librero J, Marín M, Peiró S, Aguiló J. Complication Screening Program: Validity and Sources of Data. *Int J Qual Health Care* 2000;12:439-42.
21. Iezzoni LI, Foley SM, Heeren T, Daley J, Duncan CC, Fisher ES, *et al.* A method for screening the quality of hospital care using administrative data: preliminary validation results. *QRB Qual Rev Bull* 1992;18:361-71.
22. Iezzoni LI, Davies RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, *et al.* Does the Complications Screening Programme flag case with process of care problems? Using explicit criteria to judge processes. *Int J Qual Health Care* 1999;4:107-18.
23. Hsia DC, Ahern CA, Ritchie BP, Moscoe LM, Krushat WM. Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. *JAMA* 1992;268:896-9.
24. Simborg DW. DRG Creep: A New Hospital-Acquired Disease. *N Engl J Med* 1981;304:1602-4.
25. HCIA-Sachs, Ernst & Young. *Hospital compliance and DRG upcoding.* Baltimore: HCIA-Sachs L.L.C., 2000.
26. Librero J, Peiró S, Ordiñana R. Análisis automatizado de la calidad del Conjunto Mínimo de Datos Básicos. Implicaciones para los sistemas de ajuste de riesgos. *Gac Sanit* 1998;12:9-21.
27. Guilbert A, Perez López JJ, Almela V, Company V. Calidad de datos y grupos relacionados con el diagnóstico. *Rev Calidad Asistencial* 1995;5:287-93.
28. Renau J, Perez-Salinas I. Evaluación de la calidad en la asignación de GRDs. *Rev Calidad Asistencial* 1996;11:171-6.
29. Calore KA, Iezzoni L. Disease staging and PMCs. Can they improve DRGs? *Med Care* 1987;25:724-37.
30. Thomas JW, Ashcraft ML. Measuring severity of illness: a comparison of interrater reliability among severity methodologies. *Inquiry* 1989;26:483-92.
31. Thomas JW, Ashcraft ML. Measuring severity of illness: six severity systems and their ability to explain cost variations. *Inquiry* 1991; 28:39-55.
32. Pettengill J, Vertrees J. Reliability and validity in hospital case-mix measurement. *Health Care Financ Rev* 1982;4:101-28.
33. Blumberg MS. Risk adjusting health care outcomes: a methodologic review. *Med Care Rev* 1986;43:351-93.
34. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 1982; 143:29-36.
35. Centor RM, Schwartz JS. An evaluation of methods for estimating the area under the receiver operating characteristic (ROC) curve. *Med Decis Making* 1985;5:149-56.
36. Iezzoni LI. 'Black box' medical information systems. A technology needing assessment. *JAMA* 1991;265:3006-7.
37. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA* 2000;283:1866-74.
38. Iezzoni LI, Ash AS, Coffman GA, Moskowitz MA. Predicting in-hospital mortality. A comparison of severity measurement approaches. *Med Care* 1992;30:347-59.
39. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Ann Intern Med* 1995;123:763-70.
40. Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. *Med Decis Making* 1996; 16:348-56.
41. Sainz A, Humet C. Jornada INSALUD-Sociedad Española de Calidad Asistencial. La política de difusión de los resultados de la asistencia sanitaria. *Rev Calidad Asistencial* 2000;15:448-50.
42. Peiró S, Lorenzo S. La difusión a los ciudadanos de los resultados de la asistencia sanitaria. *Rev Calidad Asistencial* 2000; 15:391-3.