ESTUDIOS DE ECONOMÍA APLICADA VOL. 24-3, 2006. PÁGS. 777-788

Potencialidades Y Limitaciones De Las Ligas De Calidad De Los Proveedores Sanitarios

MURILLO FORT; C.(*) Y GONZÁLEZ LÓPEZ-VALCÁRCEL, BEATRIZ (**)
(*) Centre de Recerca en Economia i Salut (CRES). Universitat Pompeu Fabra. Barcelona.
(**) Departamento de Métodos Cuantitativos. Universidad de Las Palmas de Gran Canaria.
*(**)*Campus Universitario de Tarifa. 35017- Las Palmas de Gran Canaria
Tel: (34) 928451800. E-mail: carles.murillo@upf.edu - bvalcarcel@dmc.ulpgc.es

RESUMEN

Las clasificaciones de centros, servicios y profesionales de la salud vienen siendo una práctica cada vez más habitual. La confección de una lista ordenada (ranking) de agentes proveedores de servicios sanitarios tratando de mejorar la efectividad y la eficiencia, tiene repercusiones financieras y de gestión. Sin embargo, el uso de la información resultante debe realizarse con suma precaución. Las listas ordenadas de centros y servicios deben cumplir con un conjunto de requisitos relativos a la construcción de la escala de medida y determinadas condiciones estadísticas y, en cualquier caso, acompañarse de elementos estadísticos que permitan medir las diferencias entre los objetos ordenados y de instrucciones para su correcta interpretación.

Palabras Clave: Ranking; clasificación de hospitales; clasificación de centros de atención primaria.

Quality Ranking Of Health Care Providers: Potential And Limitations

ABSTRACT

Ranking of health care centres, clinical services and health professionals by quality has become an usual practice. Ranking health care providers has implications on funding and management. The purpose of ranking by quality is to improve effectiveness and efficiency. But the ranking information should be used with extreme caution. The rankings should meet some properties relative to definition of the measurement scale, as well as some statistical properties. Statistical formula should be provided along with the ranking information to test the significance of the differences among the ordered objects. Instructions to interpret health care providers rankings should be also given.

Keywords: Ranking; hospital classification; primary care centers classification

Clasificación JEL: C10, I18, H5

Articulo recibido en diciembre 2006.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref.: e-24319.

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)

1. INTRODUCCIÓN. LAS LIGAS EN SANIDAD

Cada vez con mayor frecuencia y trascendencia financiera se elaboran, publican y utilizan ordenaciones (ranking) de unidades proveedoras de atención sanitaria, sean hospitales, servicios o centros de atención primaria. Muchas de estas listas ordenadas se basan en el cálculo de un indicador cuantitativo de calidad o eficiencia, que puede ser simple (porcentaje de cobertura vacunal, tasa de complicaciones post-quirúrgicas) o bien resultar de un modelo estadístico o econométrico (por ejemplo, las medidas de eficiencia resultantes de un análisis envolvente de datos o de un modelo de frontera estocástica).

En Estados Unidos, ya desde hace años se elaboran y difunden ranking de calidad de hospitales para determinadas intervenciones, es el caso del tratamiento del infarto y de la cirugía cardíaca (Hannan y otros, 1994). Incluso se comparan los resultados de médicos de atención primaria (Parkerton et alter, 2003) y de cirujanos individuales (Green y Wintfeld, 1995), ya sea utilizando soportes para expresar los resultados con detalle identificador del nombre de los médicos o de forma anónima. Algunos de los servicios que quedaron en los últimos puestos, cerraron, el resto mejoró su tasa de mortalidad en los años sucesivos (Cutler et alter., 2004). Se demostró que la experiencia influye en la mortalidad (Wu et al., 2004).

La fiebre comparativa se difundió también por Europa. Desde septiembre de 2001, el NHS británico publica, entre otros, un ranking de los consorcios que proveen asistencia hospitalaria aguda, clasificándolos con estrellas de excelencia, entre cero y tres (NHS, 2002) según un amplio conjunto de indicadores de cumplimiento de objetivos, clínicos, de gestión, de calidad, de tiempos de espera y otros que interesan al paciente. En España, una empresa privada elabora desde 2000 un ranking de calidad y eficiencia de los hospitales, públicos y privados que voluntariamente participan, basado en la agregación de indicadores de calidad, procesos asistenciales y coste (García Eroles y otros, 2001). Es el programa Top20. A diferencia de los ranking americanos antes mencionados, que ordenan por un único indicador, como la tasa de mortalidad tras una cirugía cardiaca, el Top20 español combina varios indicadores, en un ranking multicriterio que, por serlo, presenta problemas específicos y no termina de resolver el difícil equilibrio entre "la necesidad de información comparativa y la confusión" (Peiró, 2001), pues para comparar resultados hospitalarios hay que tener un marco conceptual sólido, buenos indicadores de resultados que midan calidad, ajustar adecuadamente por factores de confusión con una metodología estadística cuidada, y presentar los resultados adecuadamente, y en España estamos lejos de alcanzar esas condiciones (Peiró, 2001). Los ranking basados en encuestas de satisfacción a pacientes, utilizados en algunos países (Zwier G y D. Clarke (1999) tienen, por otra parte, su problemática específica.

Las repercusiones prácticas son tangibles, y difieren según los resultados individuales se difundan públicamente, como en el caso del NHS o de la mayor parte de

las ligas americanas, o no, como ocurre con el Top20. Algunos servicios de cirugía cardiaca que quedaban en la cola de mortalidad o complicaciones tras intervenciones de cirugía mayor, han dejado de hacerlas. El gran público conoce a través de medios de comunicación de gran tirada, las clasificaciones de algunas de estas ligas (Foster, 2001), que se han elaborado, sin embargo, con criterios excesivamente simplistas, haciendo un uso inadecuado de la estadística (Vass, 2001) y sin el necesario ajuste por riesgos.

Los problemas metodológicos de la comparación de logros de organizaciones sanitarias basada en indicadores múltiples son parientes próximos de los que surgen en la construcción de números índices con series económicas: ¿Cómo agregar los distintos indicadores parciales?; ¿Cómo ponderar, con un criterio empírico o basándonos en estándares normativos? En este trabajo no profundizaremos en esta problemática. Nos centraremos en los ranking de indicador único de tipo cuantitativo. Las propiedades deseables de ese indicador-base del ranking son: debe ser fiable, válido, sensible, preciso, con interpretación clínica, útil para la elección informada de hospital por los pacientes y para disponer de estándares de buena práctica los profesionales; fácil de obtener, y difícil de manipular. La evidencia señala que cuando se combinan diferentes variables para la construcción de un indicador final, aparecen inconsistencias entre dichas medidas individuales. La validez de estas medidas para dar respuesta a su objetivo de mejora de la efectividad del sistema es, cuando menos, cuestionable debido a la incertidumbre de los resultados. En el citado trabajo de Parkerton y otros, por ejemplo, se plantean tres hipótesis (los médicos con una calificación excelente en una de las medidas deben comportarse por encima de la media en las demás; la ordenación final de los médicos tiende a mostrar agrupaciones al considerar las distintas medidas utilizadas; el resultado de una de las medidas debe poder pronosticarse con la información de las demás medidas) que resultan rechazadas con una amplia muestra de datos procedentes de información sobre médicos de atención primaria (las medidas utilizadas tienen que ver con la satisfacción global, la gestión de los enfermos de diabetes, las tasas de screening de cáncer y los costes ambulatorios generados por cada facultativo.

En las ordenaciones subyace un proceso anidado, con al menos dos niveles. Por ejemplo, las listas de hospitales se elaboran a partir de los resultados que éstos han conseguido para sus pacientes. La propia estructura multinivel de los indicadores hospitalarios exige que los modelos estadísticos que empleemos ajusten bien por la variabilidad individual entre pacientes, introduciendo en los modelos cuando sea necesario algunas covariables específicas de los mismos. Por ejemplo, la mortalidad de los pacientes ingresados con infarto de miocardio puede depender más del estado de salud en que llegan al hospital que de la calidad de la atención sanitaria que éste les presta.

La validez de las comparaciones de calidad implícitas en las listas ordenadas de proveedores sanitarios ha sido discutida por dos tipos de problemas, en primer lugar

los asociados a la propia medida de calidad y su ajuste por factores de confusión, es decir, problemas de posibles sesgos sistemáticos relacionados con el modelo estadístico o econométrico que generó el ranking y los ajustes por riesgo, y en segundo lugar por problemas de significación estadística de las diferencias en las posiciones dentro del ranking. Puesto que los indicadores medidos son al fin y al cabo extracciones aleatorias de una distribución de probabilidad, hay errores de muestreo que además varían entre unidades. Puede ocurrir que la incertidumbre implícita en los datos sea tan elevada que de hecho no haya diferencias significativas entre centros que ocupan distintas posiciones en el ranking.

En este trabajo se discute la fiabilidad de las listas ordenadas de hospitales, la significación de las diferencias y la construcción de intervalos de confianza. El apartado 2 describe las fases en la construcción de un ranking de centros de indicador único. En el apartado 3 se discuten las cuestiones de interés estadístico relacionadas con la construcción del ranking. El artículo termina con una breve síntesis de las ventajas e inconvenientes de las ligas de hospitales.

2. FASES DEL PROCESO DE CONSTRUCCIÓN DE UN RANKING DE CENTROS

El proceso de construcción de un ranking de indicador único de centros, que también podría servir para ordenar servicios o personal sanitario, discurre por las siguientes fases:

1) Seleccionar el indicador (por ejemplo, la tasa de mortalidad intrahospitalaria tras un infarto o la satisfacción de los pacientes atendidos) y el periodo de referencia (por ejemplo, el último año)

2) Calcular el indicador bruto para cada centro, servicio o persona evaluada (i) con los datos registrados de sus n_i pacientes, a partir de una muestra representativa de los pacientes atendidos en el periodo de referencia.

3) Estandarizar el indicador. El proceso de estandarización (ajuste por riesgo, por gravedad del paciente y por condiciones del entorno del hospital independientes de su práctica o, en el caso de centros de atención primaria, por las características personales del médico sujeto a valoración) requiere un modelo estadístico de ajuste que a su vez introduce incertidumbre adicional, por los errores de medida, de especificación y de estimación

4) Ordenar los centros, servicios o personas según el indicador estandarizado, calcular los intervalos de confianza y contrastar la significación de las diferencias.

Cada una de estas fases es susceptible de problemas estadísticos que pueden invalidar los resultados. El siguiente apartado discute estos problemas. Una cuestión previa consiste en la formulación de supuestos para el comportamiento de los indicadores

calculados. Las hipótesis se formulan, al estilo de lo que realizan Parkerton y otros (op.cit.), de tal forma que su cumplimiento incorpora mayores dosis de confiabilidad en los resultados alcanzados con una aplicación práctica.

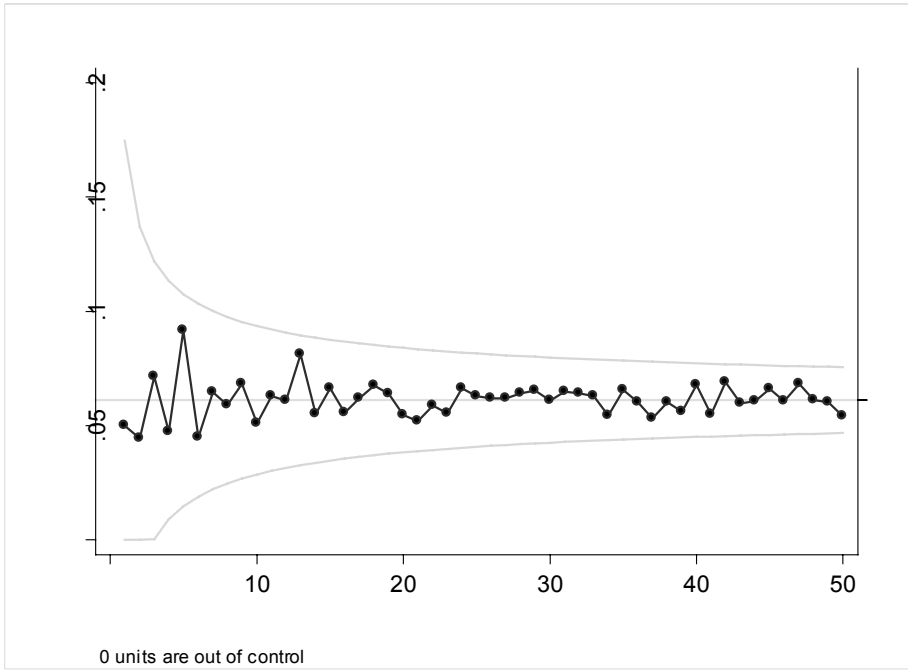
3. PROBLEMAS DE INTERÉS ESTADÍSTICO EN LA CONSTRUCCIÓN DE LOS RANKING Y DISEMINACIÓN DE RESULTADOS DE LAS LIGAS

3.1. ¿Qué queremos comparar? Método de cálculo de los intervalos de confianza y significación de las diferencias.

a) Según el objetivo de las comparaciones, habrá que elegir el método adecuado de contraste de la significación de las diferencias en el valor del indicador. Concretamente, si queremos hacer comparaciones múltiples debemos construir intervalos de confianza múltiples o simultáneos, para lo cual puede usarse un método basado en la desigualdad de Bonferroni

b) Otro enfoque clásico es el de los gráficos p, o de fracción defectuosa, que se usan en el control estadístico de calidad de procesos. La diferencia básica es que estos últimos construyen los intervalos de confianza centrados en la tasa p "estandar", única para todos los hospitales, basada en la muestra total, en datos históricos, o en estándares prefijados. La amplitud del intervalo de confianza, pues, solo depende del tamaño del centro. En el gráfico siguiente, se presenta el gráfico p para una muestra de datos simulados de 50 hospitales, ordenados por actividad (entre 40 y 2490 casos tratados, con gradiente de 50 casos) con tasas de mortalidad homogéneas del 6%.

Gráfico 1

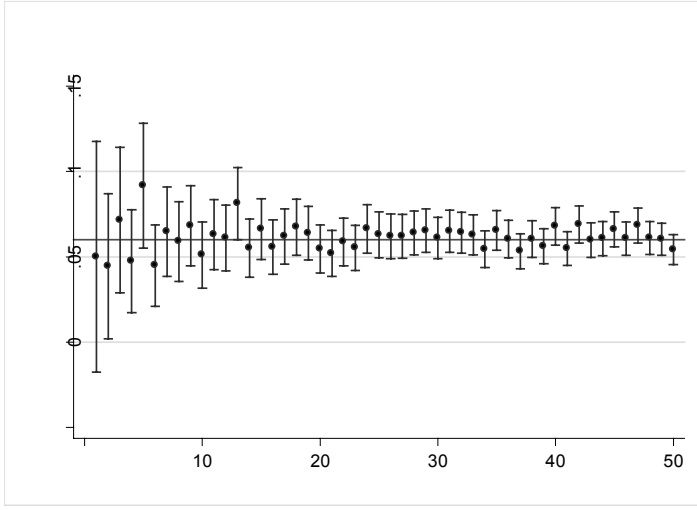


3.2. Problemas de tamaño. Efectos tamaño e incertidumbre

Si suponemos que el número de pacientes intervenidos (n_i) en cada hospital (i) es un parámetro fijo, el número de fallecidos en cirugía cardiaca es Binomial $B(n_i,p_i)$. El parámetro de interés es pi (tasa de fracasos, por mil intervenciones). Su intervalo de confianza dependerá del número de casos (n_i) y de la probabilidad de éxito y de fracaso (p_i y $1-p_i$).

El tamaño relativo de los centros (número de casos tratados) será un elemento decisivo en la fiabilidad de las comparaciones. Así por ejemplo, la distribución de una proporción p (tasa de fracasos quirúrgicos, por ejemplo) de tipo binomial, tiene por varianza p.(1-p)/n, inversamente proporcional al número de casos tratados. Los centros pequeños tendrán varianzas mayores y por tanto tenderán a situarse con mayor probabilidad en las posiciones extremas de la tabla. Este hecho se ilustra en el gráfico 2, que exhibe los intervalos de confianza del 95% para la tasa de mortalidad de los 50 hospitales anteriores. Los centros mas pequeños ocupan posiciones extremas en la liga y sus intervalos de confianza son mucho mas amplios que los de los grandes hospitales.

Gráfico 2



3.3. Problemas de ajuste y "comparaciones ajustadas"

El ajuste por riesgo es tan esencial, que un ranking que no lo haga es no solo inútil sino engañoso y contraproducente. En efecto, puede haber un problema de endogeneidad de n_i (selección por riesgo). La tasa de mortalidad no refleja realmente la calidad del servicio sino su prudencia, al admitir solo a los casos mas leves y con mayor probabilidad de éxito. La tasa de mortalidad es entonces una media ponderada de las J probabilidades procedentes de distintas distribuciones binomiales, según grupos de gravedad, y el hospital decide el número de veces que hace cada uno de los J tipos de experimento. Un problema añadido como consecuencia de las ligas es la mortalidad que sobreviene por omisión, cuando nadie acepta operar a los pacientes mas viejos o de mayor riesgo por evitar caer posiciones en la liga. La reasignación ineficiente de recursos que se produce como consecuencia es evidente, desde los más graves hacia los más leves, y el déficit de provisión de intervenciones de alto riesgo.

La "depuración" de los datos brutos para obtener un indicador estandarizado o ajustado por la gravedad inicial de los pacientes y otras condiciones independientes del hospital (ajuste por riesgo) requiere un modelo que puede ser discutible, tanto en su especificación como en el método de estimación. En cualquier caso, la variabilidad entre hospitales se estrecha tras la estandarización. Por ejemplo, un estudio comparativo de la tasa de mortalidad hospitalaria en 183 hospitales ingleses (Jarman y otros, 1999) encuentra tasas crudas de mortalidad entre el 3.4% y el 13.6%, esta variabilidad se reduce a tasas estandarizadas de mortalidad entre 53 y 137 (sobre una

media para todo el país de 100). Este indicador, tasa de mortalidad estandarizada, se basa, en definitiva, en los residuos de una regresión, que aproximan la sobre-mortalidad no justificable por las condiciones en que opera el hospital y la gravedad basal de los pacientes que admite. La pregunta es ¿hasta qué punto el modelo condiciona los datos del indicador? Goldstein y Spiegelhalter (1996) ilustran algunos casos en este sentido. Este efecto se mezcla con el efecto del tamaño, que se discute mas adelante. Hospitales mas grandes y por tanto con mayor experiencia tienden a admitir los casos mas complicados o de mayor riesgo y en este sentido presentarán mayores tasas brutas de mortalidad. Un método de ajuste por riesgo, severidad o gravedad es esencial para que las ligas sean creíbles. La regresión logística de la variable dicotómica de calidad (mortalidad, complicaciones, reingresos), estimada con datos individuales de los pacientes, es el método estadístico más usual de ajuste por riesgo.

3.4. Problemas de datos

La precariedad de los datos y los errores en las variables pueden invalidar por completo los resultados. Estos errores pueden deberse a problemas de los sistemas de información estadística, pero también ocurre que los propios hospitales manipulan a su favor la información que difunden[1], o que actúan para optimizar el comportamiento del indicador de la calidad y no la calidad misma[2]. La calidad de los datos puede ser muy variable entre Centros (Wolff y Helminiak, 1996), lo que dificulta todavía más la inferencia. Vass (2001) presenta datos de mortalidad estandarizada tras cirugía cardiaca en 29 hospitales del NHS inglés procedentes de dos fuentes diferentes. Las disparidades entre ambas fuentes son muy notables.

3.5. Problemas de difusión de los resultados

La metodología es compleja pero los resultados que se difunden se reducen a una tabla, y sólo en letra pequeña se aclaran las limitaciones metodológicas. Esta simplificación induce interpretaciones erróneas sobre la significatividad de las diferencias en los puestos de la clasificación y sobre las distancias entre las posiciones de los centros.

1 El fenómeno de codificar en un grupo diagnóstico equivocado de mayor gravedad (y financiación) es una práctica relativamente frecuente, que se conoce en la literatura como "GRD creeping"

2 Un ejemplo son las altas precoces tras un infarto, en algunos hospitales de California, para reducir la mortalidad hospitalaria. De ahí que se emplee alternativamente como indicador la tasa de mortalidad antes de los 30 días

Se ha sugerido (Tekkis y otros, 2003) una forma de informar los resultados de la mortalidad hospitalaria que supera los inconvenientes antes señalados, con un gran potencial de uso en el audit clínico (Bernal, 2003). Es una variante de los gráficos de control de procesos que incorpora tanto las tasas brutas como las ajustadas, el tamaño (número de casos) y los intervalos de confianza en torno a la tasa media global común.

4. DISCUSIÓN Y SÍNTESIS

El objetivo de las clasificaciones de centros, servicios o profesionales es múltiple. Subyace en la mayor parte de las experiencias una clara intención de mejora de los servicios, tal como aparece en la literatura publicada en el ámbito de la educación y también en el sector de la salud. Las ordenaciones atañen a centros, departamentos, servicios o personas. Consumidores, agencias de compra de servicios, aseguradoras, gestores y financiadores muestran interés por la confección de listas ordenadas de los agentes proveedores de los bienes y servicios de salud. La credibilidad de estas clasificaciones exige de los autores de las mismas una triple exigencia. Deben cumplirse, en primer lugar, ciertos preceptos de garantía metodológica en cuanto a la construcción de una escala; los autores de las clasificaciones han de garantizar la veracidad de los datos y evitar conductas oportunistas para salir mejor colocado en la lista y, finalmente, resulta de todos modos recomendable una buena guía de interpretación de los resultados que podría acompañar a las condiciones técnicas desarrolladas en el trabajo a modo, por ejemplo, de la ficha técnica que acompaña a cualquier resultado de las encuestas de opinión que los lectores consultamos periódicamente en la prensa diaria. Solamente a partir de estas premisas de seriedad y rigor metodológico la pervivencia de las clasificaciones ordenadas de centros, servicios y profesionales cumplirá con los objetivos propuestos por sus promotores.

En este trabajo se ilustra toda esta problemática con un protocolo simple de actuación en cuanto a la construcción de listas ordenadas y con un conjunto de advertencias en cuanto al tamaño de los objetos clasificados, la incertidumbre asociada a los resultados, la necesidad de ilustrar los mismos con los intervalos de confianza para cada puntuación en la escala propuesta, los errores de medida fruto de la escala de medida de las variables que recogen la información original y la exigencia de depuración de las bases de datos de valores anómalos y de ajustes para evitar efectos de confusión indeseables. Finalmente, se advierte de la conveniencia de explicar a los usuarios de la información resultante, con todo el detalle posible, del alcance de la clasificación ordenada y cómo deben interpretarse los resultados alcanzados.

5. REFERENCIAS BIBLIOGRÁFICAS

ANDERSSON, J., K. CARLING Y S. MATTSON (1998), Random Ranking of Hospital Is Unsound. Chance, vol.11 n.3 (Summer), pp. 34-39

BERNAL, E (2003) Información más creíble, con muy poco esfuerzo, ayuda a mejorar la calidad" Gestión Clínica y Sanitariae, Vol 5, n. 17, pág. 103.

CUTLER, D. M., HUCKMAN, R. S., LANDRUM, M. B. (2004): The role of information in medical markets an analysis of publicly reported outcomes in cardiac surgery. Cambridge, MA: National Bureau of Economic Research

FOSTER (2001,a), Hospital Consultants' Guide. Times supplement part I; 19 Nov, pp.22-23.

FOSTER (2001,b). Good birth guide. Times , July 15.

GARCÌA-EROLES L, ARIAS A, CASAS M. (2001) Los Top 20 2000: objetivos, ventajas y limitaciones del método. Rev Calidad Asistencial;16(2), pp. 107-16

GOLDSTEIN, H. Y D.J.SPIEGELHALTER (1996), League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. J.R.Statist.Soc. Series A 159 Part 3, pp. 385-443

HOWELL J, YONAN N, DUNN PM, ASLAN T (2002) Performance league tables. League tables are unreasonably simple, British Medical Journal, Mar 2; 324 (7336); pág. :542

JARMAN, B Y OTROS (1999) Explaining differences in English hospital death rates using routinely collected data. British Medical Journal 18, pp. 1515-20

MENEU R (2001) Top be or not Top be. Rev Calidad Asistencial;16, pp. 83-85

NHS PERFORMANCE RATINGS ACUTE TRUSTS, SPECIALIST TRUSTS, AMBULANCE TRUSTS, MENTAL HEALTH TRUSTS (2001/02).

PARKERTON, PH, SMITH, DG, BLEIN, TR, FELDBAU, GA. PHYSICIAN PERFORMANCE ASSESSMENT. NONEQUIVALENCE OF PRIMARY CARE MEASURES. MEDICAL CARE (2003); 41; 9; pp. 1034-47.

PEIRÓ, S (2001) Los mejores hospitales. Entre la necesidad de información comparativa y la confusión" Revista Calidad Asistencial, 16, pp. 119-130

RAO JN. (2001) Hospital league tables. British Medical Journal, pp.J 322: 992

TEKKIS PP, MCCULLOCH P, STEGER AC, BENJAMIN IS, POLONIECKI JD. (2003) Mortality control charts for comparing performance of surgical units. Validation study using hospital mortality. British Medical Journal, 326, pp. 786-90

WOLFF, N Y TW HELMINIAK (1996) Nonsampling measurement error in administrative data: Implications for economic evaluations. Health Economics, Vol. 5, n 6, pp. 501-512

WU, C., HANNAN, E. L., RYAN, T. J., BENNETT, E., CULLIFORD, A. T., GOLD, J. P., ISOM, O. W., JONES, R. H., MCNEIL, B., ROSE, E. A., Y SUBRAMANIAN, V. A. (2004): Is the impact of hospital and surgeon volumes on the in-hospital mortality rate for coronary artery bypass graft surgery limited to patients at high risk? Circulation, 110(7), pp. 784-789.

VASS A (2001), Doctors urge caution in interpretation of league tables. British Medical Journal, 323, pág. 1205

ZWIER G Y D. CLARKE (1999), How well do we Monitor Patient Satisfaction? Problems with the Nation-wide Patient Survey, New Zealand Medical Journal, Vol 112, pp. 371-375