

14.4 Medición y comparación de resultados en la práctica asistencial



La investigación en resultados (outcomes research) evalúa las consecuencias de las decisiones e intervenciones sanitarias en términos de su producto final: la salud individual y poblacional.

Autor: Salvador Peiró Moreno

Fundación Instituto de Investigación en Servicios de Salud, Valencia

Se recomienda imprimir 2 páginas por hoja

Citación recomendada:

Peiro, S. Medición y comparación de resultados en la práctica asistencial [Internet]. Madrid: Escuela Nacional de Sanidad; 2014 **[Actualizado abril 2023]**. Tema 14.4. Disponible en: dirección url del pdf.



TEXTOS DE ADMINISTRACIÓN SANITARIA Y GESTIÓN CLÍNICA
by UNED Y ESCUELA NACIONAL DE SANIDAD
is licensed under a Creative Commons
Reconocimiento- No comercial-Sin obra Derivada
3.0 Unported License.



Resumen:

La evaluación de servicios sanitarios puede incluir la valoración de la estructura (acreditación, certificación), los procesos (audit), los resultados o las relaciones entre estos elementos (análisis de productividad, evaluación económica, análisis de eficacia o efectividad). El análisis de resultados se emplea para evaluar la eficacia o efectividad de las tecnologías (ensayos clínicos, estudios observacionales) o para comparar organizaciones sanitarias (indicadores de actividad y perfiles de práctica). Las comparaciones entre proveedores deben ajustar las diferencias en la grave-

dad de los pacientes tratados por cada proveedor.

INDICE

1. *Introducción*
2. *Marco conceptual para la medición de resultados*
 - 2.1. *La evaluación en atención sanitaria*
 - 2.2. *Tipos de resultados*
3. *Utilidad de la medición de resultados*
4. *Estudio especial de las medidas de satisfacción, estado funcional y calidad de vida*
 - 4.1. *La medida de la satisfacción*
 - 4.2. *La medida del estado funcional*
 - 4.3. *La medida de la calidad de vida relacionada con la salud*
5. *De la evaluación a la gestión: operativizando las medidas de resultado*
6. *Referencias bibliográficas*

Las medidas de satisfacción, estado funcional y calidad de vida relacionada con la salud son posibles medidas de resultados, con ventajas e inconvenientes frente a otras medidas clásicas (mortalidad, complicaciones, etc.). Los aspectos clinimétricos (fiabilidad, validez, sensibilidad a los cambios) son esenciales en la elección de una u otra medida.

Los indicadores de actividad son la forma de operativizar las medidas de proceso y resultados en gestión sanitaria. Su utilización requiere determinados requisitos, entre los que relevancia, capacidad de ser atribuidos a los proveedores y su racionalidad

clínica son los más importantes.

IIntroducción

Probablemente los elementos que mejor configuran el entorno actual de la atención sanitaria incluyen:

- una mayor preocupación por el gasto sanitario de los gobiernos y las sociedades,
- la conciencia de amplias variaciones en la práctica médica¹,

La investigación en resultados pretende evaluar las consecuencias de las decisiones e intervenciones sanitarias en términos de su producto final, es decir la salud individual y poblacional y la satisfacción de los usuarios.

no explicables por diferencias en la morbilidad de las poblaciones,

- la preocupación por la eficacia, efectividad y seguridad de las prácticas sanitarias, que aumenta en presencia de una fortísima innovación tecnológica –en productos, antes que en procesos– y que, sin embargo, tiene un impacto decreciente sobre la salud de las poblaciones, y
- El creciente interés de los pacientes en obtener información sanitaria y en participar en las decisiones que afectan a su salud.

Este contexto ha favorecido la popularidad de una variante de investigación clínica, **la investigación en resultados de salud** (outcomes research), y la rápida difusión de muchos de sus elementos a la práctica clínica habitual.

La investigación en resultados pretende evaluar las consecuencias de las decisiones e intervenciones sanitarias en términos de su producto final, es decir, la salud individual y poblacional y la satisfacción de los usuarios.

Comparar resultados es la forma usual de evaluar la eficacia de los medicamentos, pruebas diagnósticas y otras tecnologías o intervenciones sanitarias (preventivas, educativas, organizativas, ...). Típicamente esta comparación se realizaría mediante ensayos clínicos con asignación aleatoria de pacientes a dos o más grupos. La idea básica es que las diferencias –estadísticamente significativas– en los resultados para los pacientes habrían sido causadas por diferencias en la eficacia de los tratamientos, tecnologías o intervenciones.

La comparación de resultados de fármacos o tecnologías en la práctica real (efectividad) y las comparaciones entre organizaciones sanitarias (calidad) fueron áreas de investigación minusvaloradas hasta que el desarrollo de los Patient Outcomes Research Teams (PORTs)² y la publicación de las tasas de mortalidad para más de 5000 hospitales por la Health Care

Financing Administration (HCFA) de Estados Unidos despertaron una atención sin precedentes hacia la medición de resultados en condiciones reales³.

Pese a determinadas limitaciones, el análisis de resultados ofrece oportunidades para valorar la efectividad y eficiencia de las prácticas médicas, tanto en los aspectos de evaluación de tecnologías médicas como de evaluación de organizaciones y centros asistenciales, sean estos hospitales u otros centros sanitarios, servicios o médicos individuales⁴.

En este artículo se orientan las bases conceptuales para la medición y comparación de resultados, con especial atención a sus aplicaciones prácticas. Primero se presentará un marco conceptual para la evaluación de la atención sanitaria, en el que se sitúa la evaluación de resultados, y se describen los tipos de resultados más frecuentemente utilizados. A continuación, se abordará el razonamiento básico de la utilización de las medidas de resultado en atención de salud, tanto para evaluar tecnologías como proveedores. Seguidamente, se revisarán algunos tipos de medida de resultados (satisfacción, estado funcional, calidad de vida) ya que, mientras los resultados clásicos (mortalidad, morbilidad, complicaciones, etc.) están bien aceptados en la atención clínica, los resultados que dependen de escalas de medición son menos conocidos. Finalmente, se abordará la operativización práctica de las medidas de resultado (y otras) en indicadores útiles en gestión clínica y sanitaria.

2. Marco conceptual para la medición de resultados

Durante décadas el interés en los datos de mortalidad estuvo muy vinculado al mundo académico y de investigación. La publicación por la Health Care Financing Administration (HCFA) en 1986 de las tasas de mortalidad hospitalaria de alrededor de 5000 hospitales estadounidenses provocó un interés sin precedentes hacia el análisis e interpretación de las tasas de

mortalidad hospitalarias³. La HCFA sólo mantuvo la publicación de estos rankings, con importantes mejoras metodológicas, hasta mediados de la década de los 90, pero este tipo de comparaciones y rankings no han dejado de publicarse desde entonces, incluso en España⁵

Inicialmente la mayoría de los esfuerzos se concentraron sobre las tasas ajustadas de mortalidad, pero posteriormente, y coincidiendo con el desarrollo del llamado movimiento para la medición de resultados (outcomes movement) y otras iniciativas de mejora de calidad en la práctica real, se diversificaron las medidas de resultado para detectar potenciales problemas de calidad, siendo especialmente significativos los trabajos que miden fenómenos como el reingreso, las complicaciones, el estado funcional, la calidad de vida.

La indagación en la relación entre procesos (la atención sanitaria prestada) y resultados (los cambios en el estado de salud de los pacientes) ha sido compleja y problemática. Entre las prácticas médicas aceptadas (y con excepción de los medicamentos), pocas están sólidamente fundadas en estudios bien controlados a partir de los cuales quede establecida una relación causal entre el proceso de cuidados y el resultado clínico. Esto plantea el problema de la atribución de los resultados, que, Donabedian explicó del siguiente modo⁶: “la presencia de un lazo causal válido entre un proceso específico y un resultado significa que sólo es posible alcanzar cierto resultado bajo condiciones específicas. No quiere decir que el resultado observado en una situación dada ha sido producido, en ese caso, por el proceso citado (...)”.

En otras palabras, cuando se usan resultados para hacer inferencias acerca de la calidad de los cuidados, es necesario establecer qué resultados pueden ser atribuidos a esos cuidados (validez de atribución). Hay que resaltar que esta clase de validez depende del establecimiento de una relación causal entre procesos y resultados sobre bases científicas y es –todavía hoy día– el principal problema para la aplicación práctica de la

evaluación de resultados a la atención sanitaria.

2.1. La evaluación en atención sanitaria

Donabedian conceptualizó los ejercicios de medición-evaluación de la atención sanitaria en la tríada (**Figura 1**) estructura, proceso y resultados⁷.

- La **estructura** incluye los soportes físicos, económicos, de personal, de formación, organización, etc., destinados a la atención de salud, incluyendo los conocimientos existentes y las capacidades y habilidades del personal.
- Los **procesos** se refieren a aquello que los proveedores de servicios sanitarios (hospitales, centros de salud, médicos, etc.) hacen a, por y para los pacientes, y también como los pacientes buscan la atención médica y responden a las pautas terapéuticas recomendadas.
- El impacto de estos procesos sobre la salud de los pacientes son los **resultados o desenlaces** (outcomes, endpoints), definidos como los **cambios, favorables o no, en el estado de salud actual o potencial de personas, grupos o comunidades, que pueden ser atribuidos a la atención sanitaria previa o actual.**

Figura 1.

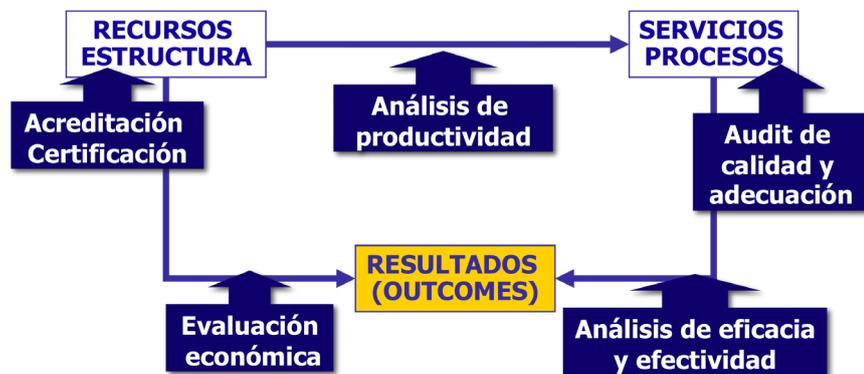
Marco conceptual de la medición-evaluación de servicios sanitarios: modelo clásico



Los indicadores pueden interpretarse mediante normas estadísticas, normas fijadas por autoridades sanitarias o profesionales y normas basadas en evidencias científicas.

Sobre este marco es posible situar las diversas modalidades de evaluación que corresponderían a diversos tipos de medidas o indicadores (**Figura 2**). Las **medidas de estructura**, aunque a veces se utilicen por su fácil disponibilidad, tienen escaso interés desde la perspectiva de la evaluación de resultados. En parte porque las más importantes suelen exigirse como requisito en los procedimientos de autorización previa de centros sanitarios, en parte por su escasa correlación –al menos en el entorno de los países con sistemas hospitalarios modernos– con los resultados para los pacientes.

Figura 2.
Marco conceptual: modelo clásico



Las medidas de proceso, incluyendo las de utilización de servicios, pueden incluir:

- Sucesos deseables (ej.: porcentaje de mujeres de determinados grupos de edad que acuden a los programas de cribado de cáncer de mama),
- Sucesos esperables por la condición de los pacientes pero que en cifras elevadas (o bajas) sugieren problemas de calidad (ej.: porcentaje de cesáreas), o
- Sucesos no deseables (ej.: porcentaje de pacientes con más de 2 meses de espera entre un cribado positivo y la cirugía de mama, hipertensos no controlados, reingresos inesperados, etcétera).

Los indicadores, usualmente en tasas o porcentajes respecto al denominador de interés, pueden interpretarse mediante normas estadísticas, normas fijadas por autoridades sanitarias o profesionales o normas basadas en evidencias científicas.

Las **normas estadísticas** se aplican a grupos y permiten valorar si los proveedores actúan de forma diferente, pero no la adecuación de la atención en pacientes individuales. Usualmente derivan de la práctica observada en un determinado conjunto de proveedores y –cuando se adopta el promedio como referencia– tienden a reforzar el statu quo por lo que se ha sugerido que las comparaciones deberían realizarse respecto a centros de excelencia en un determinado campo.

Las **normas basadas en la evidencia científica** representan criterios explícitos de cuidados apropiados basados en resultados de estudios clínicos. Usualmente toman la forma de porcentaje de cumplimentación de una recomendación normativa para la atención de una situación concreta en un grupo definido de pacientes (p. ej. trombólisis en la hora siguiente al ingreso hospitalario en pacientes con infarto agudo de miocardio y sin contraindicaciones para el fármaco; o porcentaje de estos pacientes que habían recibido ácido acetil salicílico al inicio de síntomas).

Las **normas basadas en consensos** son similares pero su base evidenciaría es más débil y responden a acuerdos entre expertos (p. ej. porcentaje de diabéticos insulín-dependientes que reciben un control de glucemia trimestralmente y una revisión oftalmológica anualmente).

Las medidas de proceso son el eje central de los indicadores de funcionamiento –mucho más que las de resultados– y, cuando los ensayos clínicos han demostrado que el proceso medido está asociado a mejores resultados, la devolución de la información a los clínicos es uno de los instrumentos más potentes de mejora de calidad.

En la relación entre recursos y resultados se sitúan las **medidas**

de productividad, entendidas como la relación entre los recursos (expresados en unidades físicas: médicos, enfermeras, camas, quirófanos, etc.) y determinados procesos que se utilizan como productos intermedios (expresados también en unidades físicas: estancias, intervenciones, pruebas diagnósticas, etc.); aunque el análisis de la productividad en los servicios de salud se enfrenta a importantes problemas conceptuales y prácticos en la definición del recurso y del producto, suele tratarse de medidas de fácil obtención que forman la parte más voluminosa de los indicadores usuales de funcionamiento hospitalario (estancia media, índice de rotación, funcionamiento de quirófanos y similares).

Las **medidas de eficiencia** se centran en la relación entre recursos (medidos ahora como unidades monetarias) y resultados de eficacia o efectividad. Por ejemplo, coste por caso de hepatitis evitada o por año de vida ganado. Su instrumental básico es la evaluación económica en cualquiera de sus modalidades, especialmente los análisis coste-efectividad y coste-utilidad. Como indicadores de funcionamiento suelen emplearse en forma de medidas indirectas derivadas de estudios que han demostrado la eficiencia de alguna práctica (p. ej. porcentaje de prescripción de genéricos, o de cirugía ambulatoria en determinadas intervenciones).

Finalmente, las **medidas de resultado** se centran en aquellos aspectos realmente importantes para los pacientes: mortalidad, complicaciones, estado funcional, calidad de vida, satisfacción, etc. Pese a su creciente utilización, **son ampliamente discutidas ya que los resultados no dependen exclusivamente de la atención prestada**. Ejemplos típicos de estos indicadores son las tasas de mortalidad tras cirugía cardíaca, las tasas de complicaciones de herida, de infecciones nosocomiales, la satisfacción, etc. Atendiendo al sentido de los cambios, los resultados pueden considerarse favorables (reflejando una mejora en una o más de las dimensiones del estado de salud) o adversos (representando un deterioro).

Dado que los procesos de cuidados afectan diferentes dimensiones

del estado de salud y que los resultados pueden interrelacionarse, es necesario distinguir entre **resultados finales** (endpoints) y **resultados intermedios** (surrogate endpoints). Los resultados finales son aquellos que tienen relevancia clínica para los pacientes y son los que realmente se desea conseguir con las intervenciones sanitarias. La reducción de la mortalidad es el clásico ejemplo de resultado final, pero también lo es la mejora en el estado funcional o en la calidad de vida. La investigación clínica a menudo utiliza resultados intermedios (por ejemplo, control de la presión arterial) como evidencia de éxito en el tratamiento o en el proceso de cuidados, pero este tipo de resultados debe ser visto con cautela, porque lo que realmente se desea es reducir la mortalidad (o la mortalidad cardiaca, o la incidencia de infartos) y no en todos los casos y en todos los grupos de pacientes la asociación entre surrogate endpoints y endpoints tiene la misma importancia.

2.2. Tipos de resultados

En la **Tabla 1** se muestran los resultados más empleados en atención de salud con algunos ejemplos de operativización de estos resultados. La **mortalidad**, uno de los resultados más utilizados por su rotundidad y fácil disponibilidad, suele emplearse en diferentes ventanas temporales en relación con el tipo de patología o con el proceso de interés (por ejemplo, los cirujanos están especialmente interesados en la mortalidad a los 10 y 30 días tras la intervención quirúrgica), por causas específicas (mortalidad neonatal, maternal, etc.) o como supervivencia tras el diagnóstico (usual en oncología). Es un indicador de poca utilidad para evaluar procesos con baja mortalidad prevista (por ejemplo, la cirugía sin ingreso), aunque en estos supuestos puede tener interés el análisis detallado de los casos concretos. Tampoco es muy útil para evaluar centros de atención primaria, y no sólo por limitaciones de tipo estadístico, sino por la dificultad de atribuir las diferencias en mortalidad a diferencias en calidad de la atención primaria.

Las **complicaciones y otros sucesos adversos**, como los reingresos, las infecciones nosocomiales y algunas complicaciones específicas (de herida quirúrgica, recidivas, fracasos de intervención), son cada vez más empleadas junto a -o en lugar de- la mortalidad por su mayor frecuencia (mayor poder estadístico) y por una presumible mayor relación con problemas de calidad en la atención susceptibles de corrección. En este terreno cabe señalar la creciente utilización de este tipo de medidas en relación con la prescripción farmacéutica (**reacciones adversas a medicamentos**). Se ha propuesto el uso de algunos tipos de complicación que suelen requerir atención hospitalaria en urgencias o bajo ingreso (por ejemplo, descompensaciones diabéticas) para evaluar la calidad de la atención primaria a partir de los -más consolidados- registros hospitalarios. Las propuestas de utilizar las "hospitalizaciones evitables" para monitorizar la calidad de la atención primaria irían en este sentido, aunque no sin limitaciones⁸.

La **estabilidad o mejoría de los sistemas fisiológicos** es empleada como medida intermedia de resultado en muchas patologías crónicas (disminución de la presión arterial en mmHg, reducción de los niveles hemáticos de lípidos, control de la glucemia o de la hemoglobina glicosilada, etc.) que suelen controlarse en atención primaria. También es empleada en las unidades de cuidados intensivos donde muchas de estas variables son críticas en la valoración de resultados inmediatos (normalización gases, frecuencia cardíaca, ...).

La medida de la **capacidad funcional** es muy empleada en especialidades como la cardiología o la neumología, con clasificaciones históricas como la de insuficiencia cardíaca de la New York Heart Association (NYHA) utilizada desde 1972, y pruebas funcionales objetivas de amplio uso como las respiratorias, de esfuerzo y otras. En otras áreas, como la cirugía ortopédica o la rehabilitación, es la forma esencial de medir resultados importantes para los pacientes. La capacidad funcional también incluye las medidas de **funcionamiento psico-social**.

Tabla 1. Resultados de la atención de salud

<i>Resultados</i>	<i>Ejemplos</i>
Supervivencia o mortalidad	Mortalidad intrahospitalaria; Mortalidad a los <i>n</i> días de la intervención; supervivencia a los <i>n</i> años del diagnóstico.
Otros sucesos adversos	Reingresos en determinados periodos o por causas específicas; infecciones nosocomiales; complicaciones post-quirúrgicas; complicaciones de la herida quirúrgica, reacciones adversas a fármacos; reintervenciones inesperadas; recidivas a los <i>n</i> años de la intervención.
Estabilidad de los sistemas fisiológicos	Normalización/mejora de signos clínicos: pulso, temperatura, u otros; normalización/mejora de parámetros bioquímicos; normalización mejora de otros parámetros.
Satisfacción	Encuestas de satisfacción y otros instrumentos
Estado funcional	Capacidad para el esfuerzo; capacidad para las actividades de la vida diaria; capacidades específicas: visuales, movilidad, y otras similares; normalización/mejora en pruebas funcionales; reincorporación a la vida laboral.
Funcionamiento psico-social	Estado cognitivo; escalas de ansiedad, depresión, estrés, y otras similares.
Dolor	Escalas analógico-visuales, instrumentos específicos de medición del dolor.
Calidad de vida relacionada con la salud	Medición de utilidades; índices de estado de salud; perfiles de salud genéricos o específicos; instrumentos diagnóstico-específicos de medición de calidad de vida.

La medida de la capacidad funcional implica la valoración por alguien **externo al paciente** y con **criterios objetivos** de los efectos de la enfermedad sobre la vida del individuo; la valoración funcional excluye los aspectos subjetivos. Las mediciones se concretan como puntuaciones en **escalas de valoración funcional genéricas** (índices de actividades de la vida diaria) **o específicas de las patologías a valorar** (por ejemplo: Knee Society Score, Harris Hip Score) o resultados de pruebas funcionales (audiometría, pruebas de esfuerzo, de agudeza visual, etc.). Lo esencial de estos instrumentos es que permitan cuantificar los resultados clínicos relevantes de un modo factible, uniforme, objetivo y reproducible.

La medición de la **calidad de vida relacionada con la salud** (HRQoL, CVRS) supone un esfuerzo para medir de forma comprensiva los resultados importantes para los pacientes. En cierto modo, surge como una extensión de las medidas de capacidad funcional a las que se incorporan una serie de

componentes, entre los que destacan la **percepción subjetiva por parte del paciente del impacto de la enfermedad** (aspecto que requería un enfoque multidimensional y no solo de función) **y la relación con los cuidados de salud**⁹⁻¹¹. Este último aspecto es trascendente porque el término calidad de vida es excesivamente general y las aproximaciones útiles en el contexto de la atención médica se concentran en aquellos aspectos de la experiencia vital de las personas relacionados con la salud y la asistencia sanitaria (aspecto que es descrito con mayor exactitud por el término calidad de vida relacionada con la salud, health related quality of life).

Finalmente, la **satisfacción de los pacientes**, acompañando a las transformaciones sociales y el cambio del papel de los usuarios de los servicios de salud, se ha ido convirtiendo en un resultado importante, para cuya medición se han desarrollado numerosos marcos conceptuales y encuestas específicas. Cabe señalar, no obstante que este tipo de encuestas se utilizan más como instrumentos de mejora de calidad que como medida de resultados.

2. Utilidad de la medición de resultados

Comparar resultados es la forma usual de evaluar la eficacia de los medicamentos, pruebas diagnósticas y otras tecnologías. Típicamente esta comparación se realiza mediante ensayos clínicos con asignación aleatoria de pacientes a dos o más grupos. La idea básica es que las diferencias 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 a 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.

El problema fundamental al comparar los resultados fuera de situaciones experimentales estriba en que, a diferencia de los ensayos clínicos, los pacientes no han sido asignados aleatoriamente a los tratamientos o al tratamiento en un u otro centro sanitario. Al menos en buena parte, han sido asignados en función de su propia condición clínica. En esta situación no puede presumirse que las diferencias en resultados se deban a exclusivamente a la calidad de los centros sanitarios, ya que estarán muy influidos por las características de los pacientes tratados en cada lugar.

Las características de los pacientes que definen su riesgo de peores resultados se agrupan bajo el nombre de **gravedad** (severity). Las características de los pacientes que configuran la gravedad incluyen factores como la edad, el sexo, el diagnóstico principal y su gravedad, 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 inicio del proceso), la estabilidad clínica, el estado funcional, el estado psicológico, cognitivo y psico-social, y otras variables.

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 en cada centro por aquellas características de los pacientes que definían su riesgo previo respecto al resultado que se utiliza³.

Si se quiere valorar diferencias en cualquier resultado (muertes, reingresos, complicaciones, calidad de vida) entre tratamientos u 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 centro 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.).

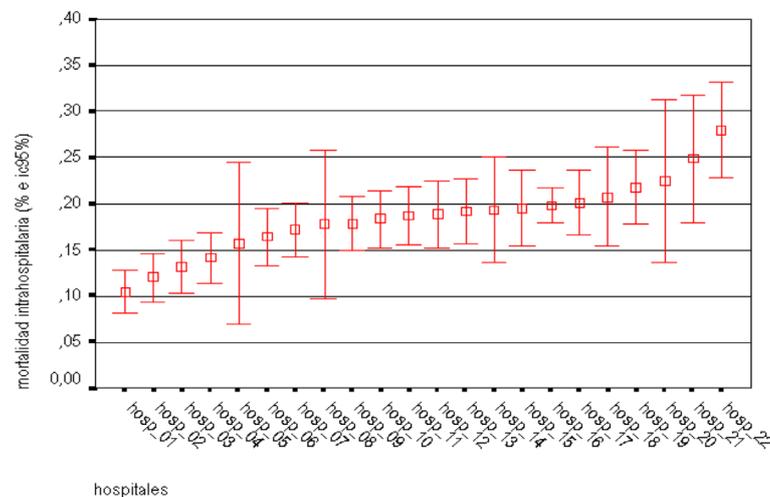
Esta es la base para la construcción de sistemas de clasificación de pacientes o -generalizando, ya que hay sistemas que no construyen grupos de pacientes- sistemas de ajuste de riesgos (risk adjustment systems) que, simplificando, no son más que métodos de cuantificar la probabilidad que tienen los pacientes de obtener un determinado resultado.

En la Figura 3 se muestran las tasas de mortalidad intrahospitalaria, incluyendo sus intervalos de confianza del 95%, en el ictus cerebral (excepto isquemia transitoria) en 22 hospitales públicos de la Comunidad Valenciana (1993-1995). Puede verse que estas tasas variaban del 10% al 28% con diferencias estadísticamente significativas entre algunos hospitales (los intervalos de confianza no se solapan).

Figura 3.

Medidas de resultado: calidad de los proveedores

**MORTALIDAD INTRAHOSPITALARIA EN EL GDR 14 (ACCIDENTE VASCULO-CEREBRAL excepto ISQUEMIA TRANSITORIA).
Hospitales públicos del País Valenciano (1993-95)**



Estos datos, aplicando el razonamiento previo de inferir desde las diferencias en resultados hacia la calidad de la atención prestada, implicarían graves problemas de calidad. En la práctica, la interpretación no es tan simple, ya que los hospitales sin servicio de neurocirugía derivaron los casos más graves (ictus hemorrágicos, con mortalidad usualmente superior al 30%) hacia los hospitales que disponían de estos servicios, mientras que retuvieron los accidentes trombóticos (con mortalidad en torno al 9%). En el caso presentado, en el hospital con la menor tasa de mortalidad sólo el 25% de las altas se debían a hemorragias cerebrales, mientras que esta cifra era del 57% en el hospital con mayor mortalidad. Lógicamente, la mortalidad fue mayor en este último, pero este exceso de mortalidad no se debió tanto a un problema de calidad como a que los "peores" pacientes fueron razonablemente derivados hacia los "mejores" centros.

Esta situación, corriente (y deseable) en la atención sanitaria práctica (como norma se intentará que los pacientes sean atendidos en aquel lugar con mayores recursos y experiencia en la patología en cuestión) hace que las comparaciones de resultados entre organizaciones sanitarias deban ser realizadas con gran cautela, sea cual sea la medida de resultado empleada e incluso cuando se emplean sistemas de ajuste de riesgos. Como norma, los resultados de este tipo de análisis (estudios observacionales, y usualmente sobre datos tomados retrospectivamente) nunca tendrán la misma solidez para inferir desde los resultados a la calidad de los centros sanitarios, que la que muestran los ensayos clínicos con asignación aleatoria para inferir desde los resultados a la eficacia de una tecnología médica concreta^{12,13}. Pese a ello, es el tipo de análisis que puede utilizarse para evaluar los resultados de los proveedores y, en la práctica y aplicado con prudencia, puede ser extremadamente útil para mejorar la calidad de la atención.

3. Estudio especial de las medidas de satisfacción, estado funcional y calidad de vida

3.1. La medida de la satisfacción

La medida de la satisfacción de los pacientes captura un tipo de información –la valoración personal de los usuarios en relación con sus preferencias y expectativas– que no puede obtenerse mediante la observación externa y, por tanto, queda más allá de la información que pudiera obtener el proveedor utilizando instrumentos objetivos. Desde los trabajos iniciales de Hulka^{14,15}, tanto los sociólogos como los investigadores en servicios de salud han desarrollado diversas teorías y modelos para definir los “puntos de vista” de los pacientes sobre la atención que reciben, sin que todavía exista una definición aceptada y una clara conceptualización de la “satisfacción” de los pacientes.

La taxonomía de componentes en relación con la experiencia de los pacientes que se incluyen en los instrumentos de medida de la satisfacción es extraordinariamente diferente y muestra una enorme variabilidad en las formas de describir la valoración de estas experiencias en los diferentes estudios (excelente o pobre, satisfactoria o insatisfactoria, agradable o desagradable, disponibilidad o no a volver o a recomendar el servicio, etc.), las tasas de respuesta, el coste y aceptabilidad para los posibles usuarios de las encuestas, las formas de evaluar la validez utilizadas –con grandes insuficiencias– y los posibles sesgos (no respuesta, forma de administración, tiempo desde el alta, formato de las respuestas, sesgo de contexto, respuesta por familiares).

Esta situación de discrepancias en la definición y conceptualización de la satisfacción de los pacientes, junto a una gran difusión de las encuestas formando parte de estrategias de marketing o mediáticas (las encuestas de satisfacción no son sólo un instrumento de investigación, sino que están incorporadas a la

gestión rutinaria de muchos centros y organizaciones sanitarias), ha propiciado la existencia de cientos, probablemente miles, de instrumentos de medida de la satisfacción diferentes cuya elección se basa más en la casualidad, la moda o aspectos coyunturales, que en el conocimiento de sus propiedades clinimétricas o la asunción de algún marco conceptual definido.

Hulka^{14,15} conceptualizó su aproximación a la **medición de la satisfacción de los pacientes como una medida de resultado**, fundamentalmente dirigida a comprender las actitudes hacia el sistema sanitario en diferentes segmentos de población. En este sentido, señalaba que la satisfacción de los pacientes debía valorarse como un resultado per se que, adicionalmente, influiría en la utilización y conducta de los pacientes (adherencia, fundamentalmente).

Ware^{16,17}, sin embargo, consideraba la satisfacción **como una evaluación personal –subjetiva– de los servicios sanitarios y de los proveedores**. Aunque reconocía que un nivel determinado de satisfacción aportaba información desde el punto de vista de las preferencias del paciente, enlazando de este modo con la perspectiva del marketing, no transformaba la definición y medida de la satisfacción en un análisis de gaps, de elementos a corregir, sino en una aproximación a la calidad de los procesos, antes que los resultados, de un sistema sanitario concreto.

Parasuman^{18,19}, desde el campo del marketing de servicios, analizó las **diferencias (gaps) entre el servicio recibido y el servicio esperado, tanto por parte de los médicos como del usuario**. Estos estudios buscaban la identificación de gaps para adaptar los servicios a las expectativas de los “clientes” en 3 áreas: expectativas del usuario frente a experiencias del usuario, expectativas del usuario frente a percepciones profesionales de las expectativas del usuario, y experiencias del usuario frente a percepciones profesionales de las experiencias del usuario. En atención sanitaria, el desarrollo más importante de esta

perspectiva es el SERVQUAL que mide la calidad de los servicios como diferencias entre las expectativas y las percepciones de los pacientes y ha sido utilizada en numerosos estudios en España.

Hay que señalar que, aunque la difusión y aplicación real de los tres modelos (Hulka, Ware, Parasuman) ha sido diferente, los tres continúan actualmente en uso. La encuesta original de Hulka se emplea poco, pero existen numerosas adaptaciones en uso, especialmente en atención primaria, y son la base de la mayor parte de las encuestas utilizadas en los hospitales. El trabajo de Ware y su equipo ha tenido una amplia difusión y aplicación en el sistema sanitario americano, dando origen al Consumer Satisfaction Survey, un instrumento bien estudiado que pretende medir la satisfacción de los usuarios con los planes de aseguramiento sanitario con el fin de que los financiadores de los servicios, básicamente las empresas que aseguran a sus trabajadores, pudieran establecer una comparación entre ellos. El SERVQUAL, por su parte, es posiblemente la escala de medición de calidad más conocida y usada en Estados Unidos y, aunque SERVQUAL no es solo un instrumento de medida de satisfacción, en atención sanitaria se ha centrado en este aspecto y en el de calidad percibida.

Una característica de los instrumentos de medida de la satisfacción de los pacientes es que **los resultados tienden a estar sesgados positivamente**. En torno al 85%-95% de los encuestados elegirá una respuesta u opinión positiva. Una explicación a esto es el efecto de la "deseabilidad social", en el que los encuestados encuentran muy difícil expresar críticas sobre la asistencia sanitaria. En todo caso, este sesgo positivo convierte las encuestas de satisfacción en un instrumento muy poco sensible a los cambios y poco útil, en la práctica, como medida de resultado que permita valorar la calidad de la atención recibida

3.2. La medida del estado funcional

El estado funcional es conceptualizado como la capacidad para procurarse los **propios cuidados y desarrollar las actividades físicas**. Hasta que surgieron las escalas de valoración funcional, en la década de los 50, la capacidad funcional apenas fue empleada como medida de resultados.

El objetivo básico de éstas escalas es permitir la valoración inicial de la función (general, de un órgano, de una articulación, etc.) tras la identificación de la patología, hacer el seguimiento del proceso, valorar los resultados una vez finalizado el proceso, comparar resultados con otros profesionales que han utilizado la misma intervención terapéutica o comparar diferentes terapias ante un mismo problema y, en resumen, proporcionar una base para la comunicación científica y para evidenciar de forma objetiva la eficacia de los tratamientos.

Se hace hincapié en el **aspecto objetivo de la valoración funcional** que, a diferencia de la CVRS, siempre es evaluable por una persona externa al paciente. No obstante, algunas escalas incluyen la valoración del dolor, aspecto en buena medida subjetivo y que las aproxima a las escalas de calidad de vida.

Las **actividades de la vida diaria** (AVD) fueron el primer parámetro funcional ampliamente aceptado y reconocido y se definen como **aquellas funciones físicas, diarias y rutinarias, necesarias para vivir**. Incluyen actividades de cuidado personal (vestido, baño, aseo personal, uso del retrete, transferencias, alimentación, continencia, movilidad en interiores) y, aunque existe un acuerdo general en lo que constituye el núcleo central de las AVD, las llamadas AVD básicas (AbVD), existen variaciones en las actividades incluidas en diferentes escalas. Las escalas de valoración de las AbVD más utilizadas son el **índice de Actividades de la Vida Diaria** (Index of Independence in Activities of Daily Living) y el **índice de Barthel**. La principal limitación de las escalas de AbVD es su escasa sensibilidad

para detectar cambios clínicos pequeños, aunque estos sean relevantes. Para corregir este aspecto se desarrollaron las escalas de valoración de las AVD instrumentales (AiVD), que incluyen actividades más complejas (necesarias para vivir independiente en el domicilio y en la comunidad: preparación de la comida, cuidado del hogar, lavado de ropa, capacidad para usar el teléfono, uso del transporte público, responsabilidad con el uso de la medicación y manejo del dinero, así como el cuidado de niños y la movilidad fuera del hogar) y son más sensibles a los cambios. Entre las escalas de AiVD más utilizadas se encuentran the Physical Self-Maintenance Scale, más conocida como **índice de Lawton**.

Aunque la capacidad para realizar las AVD se ha usado como sinónimo de función o rendimiento físico, sólo supone una parte de la función física global que abarcaría también las dimensiones mental, emocional y social. Dentro de este contexto se han elaborado escalas de valoración funcional globales o mixtas que incluyen aspectos de la función física, de la comunicación, cognitivos, sociales y psicosociales, entre las que destaca el **Functional Status Index**.

Igualmente, muchas escalas son específicas de un órgano (por ejemplo: escalas de agudeza visual) o de una articulación (por ejemplo, la Escala de Cadena de Harris) y, en algunos casos, los resultados de algunas pruebas son equiparables a la medida de la función de un órgano (por ejemplo: audiometría, pruebas funcionales respiratorias, ergometría, ...). Señalar, finalmente, que las escalas de valoración globales incluyen aspectos sobre los que no es posible la intervención médica o de enfermería, por lo que, en la práctica, en algunos casos se seleccionan para uso sólo aquellos aspectos de una escala más relevantes para la evaluación de un proceso concreto.

3.3. La medida de la calidad de vida relacionada con la salud

La calidad de vida es un concepto fácil de comprender intuitivamente, pero complejo de definir y sobre el que no existe un acuerdo universal. En medicina se suele emplear el concepto de calidad de vida relacionada con la salud (health related quality of life) que mediría **el impacto de la enfermedad y sus tratamientos sobre el bienestar del paciente, incorporando –más allá de las mediciones clínicas- la perspectiva del propio paciente**^{20,21}.

La CVRS es un concepto subjetivo que sólo puede ser valorado por el propio paciente. La valoración por los profesionales no sólo es inadecuada sino, como han demostrado varios trabajos en los que estas valoraciones diferían ampliamente de las realizadas por los pacientes, inexacta. Hay que señalar que la subjetividad de las medidas de CVRS no implica falta de reproducibilidad que, en general, es similar a la de las medidas facilitadas por instrumentos objetivos.

La CVRS es un concepto multidimensional (para incorporar los distintos aspectos del bienestar), subjetivo (para recoger la perspectiva del paciente) y dinámico (cambia a lo largo del tiempo, pero también en función del entorno).

Las primeras medidas de calidad de vida por su proximidad -cuando no confusión- con las medidas de capacidad funcional enfatizaban los componentes relacionados con la función física, pero la consideración multidimensional de la CVRS obliga a identificar y operativizar los componentes que la van a integrar. Estos componentes son denominados dimensiones o dominios y, como mínimo, incluyen la dimensión física, la psicológica y la social. Aunque inicialmente las medidas de CVRS se apoyaban sobre todo en la dimensión física, evolutivamente se han ido incluyendo un mayor número de dimensiones pertenecientes a la vida psicológica y social, y se ha dado una mayor especificación

los dominios que conforman cada una de ellas.

Las dimensiones de mayor interés podrían describirse como:

1. **Función física.** Similar al concepto de capacidad funcional. Se refiere a la energía, fuerza, y capacidad para realizar una actividad usual. contempla fundamentalmente 2 dominios:
1) La capacidad para las **actividades de la vida diaria** y,
2) La **capacidad para el esfuerzo**, o energía o actividad positiva.
2. **Función psicológica.** Se refiere a indicadores cognitivos y afectivos. **La función afectiva**, referida al estrés psicológico ha producido el desarrollo de numerosas escalas orientadas sobre todo a la evaluación de la depresión y ansiedad. La **función cognitiva** es también considerada un importante componente del bienestar psicológico, ya que la memoria y la capacidad para llevar a cabo tareas intelectuales tienen una gran importancia para las personas.
3. **Interacción y función social.** Se refiere a la capacidad del paciente para desarrollar las interacciones personales que forman el núcleo de la vida social. Estas interacciones están tradicionalmente estructuradas en una forma jerárquica: familia, amigos, compañeros y comunidad.
4. **Sensaciones somáticas, limitaciones y alteraciones.** Reúne las características de orden fisiológico derivadas de la enfermedad y/o tratamiento. En general se consideran los síntomas o molestias referenciadas por el paciente, las sensaciones displacenteras (dolor, náuseas y disnea entre otros) y los signos obtenidos de la exploración clínica, los resultados o medidas fisiológicas provenientes de pruebas diagnósticas. Son fundamentales en los instrumentos de calidad de vida diseñados para enfermedades específicas.
5. **Percepciones de salud.** Basadas en las propias percepciones del paciente, incluyen las evaluaciones que

el individuo hace de los efectos que tiene su salud sobre otros aspectos de la vida.

Pese a la importancia de la conceptualización de las dimensiones incluidas en cada instrumento de medición, en la actualidad sigue sin existir uniformidad en la cantidad, significados, e importancia con la que contribuyen en la calidad de vida. Por otra parte, la influencia de valores socioculturales en la calidad de vida percibida corrobora la necesidad de utilizar componentes acordes al contexto a evaluar.

La información sobre calidad de vida puede ser obtenida por medio de entrevistas y preguntas abiertas y/o cerradas. Esta información puede ser muy útil para el manejo de un paciente individual, pero salvo que pueda ser cuantificada en alguna medida, su utilidad para transmitir información científica es pequeña. Por este motivo, la información sobre CVRS es obtenida mediante cuestionarios estructurados que incluyen ítems (preguntas o afirmaciones que pueden puntuarse dicotómicamente o en escalas ordinales o categóricas) para recoger las diferentes dimensiones. Las respuestas a estas preguntas se cuantifican en puntuaciones para cada una de las dimensiones y/o en una sola puntuación global.

Los instrumentos de medida de la CVRS pueden clasificarse según diferentes criterios como el tipo de población a la que se dirige, los dominios y/o conceptos que cubre, el sistema de puntuación bajo el que refleja el nivel individual y/o global de los componentes y algún otro. No obstante, para diferenciar el conjunto de medidas que incluyen diversos aspectos relacionados con la calidad de vida, el criterio más utilizado es el que se refiere a su aplicación general (a diferentes poblaciones) o específica (enfermedades concretas), conocidas comúnmente como **instrumentos genéricos e instrumentos específicos**.

El propósito de los **instrumentos genéricos** es ser aplicables a una gran variedad de poblaciones. Se dividen perfiles e índices de utilidad. Los **perfiles de salud** miden diferentes aspectos de

la calidad de vida utilizando un sistema de puntuación común, que puede corresponder a un número reducido de ítem que conforman una dimensión, o bien algunas veces también permiten obtener una puntuación global, a partir de la suma de valores alcanzados en cada dimensión. La importancia y el número de ítem otorgados a cada dominio puede ser muy diferente según perfiles. Así, existen instrumentos que exploran las dimensiones físicas, emocionales y sociales de modo equitativo mientras otros cargan sobre algún dominio un número mayor de ítem, enfatizando con ello su valoración.

Como principales ventajas, los perfiles suelen tener estudios de fiabilidad y validez bien establecidos, con frecuencia dirigidos a distintas poblaciones. Para propósitos discriminativos, se puede examinar y determinar el área de disfunción que afecta a un individuo o grupo. La identificación de estas áreas permite a su vez, guiar la elaboración de instrumentos específicos y orientar la intervención clínica hacia donde produzca un mayor impacto en la calidad de vida. Cuando se aplican con propósitos evaluativos, también permiten determinar los efectos de la intervención sobre diferentes aspectos de la calidad de vida sin necesidad de aplicar múltiples instrumentos. Por otra parte, facilitan la realización de estudios comparativos en distintas poblaciones. Por último, aquellos que proporcionan una puntuación global, pueden utilizarse en los análisis de coste-beneficio sobre las intervenciones clínicas. Entre las desventajas, pueden focalizar su atención hacia ciertos aspectos de la calidad de vida que no sean los que más interesan al objetivo de un determinado estudio, perdiendo sensibilidad cuando se aplican a poblaciones con enfermedades específicas.

Los perfiles de salud más conocidos incluyen **el Medical Outcomes Study Short Form (MOS-SF)**, que en sus diversas versiones con más o menos ítem (SF-36, SF-12, SF-8) es probablemente el instrumento más utilizado en la actualidad, **el Perfil de Salud de Nottingham** (Nottingham Health Profile, NHP), el Perfil de Impacto de la Enfermedad (Sickness

Impact Profile, SIP) y **las láminas COOP-WONCA**, de uso fundamentalmente en atención primaria. Todos ellos disponen de numerosos estudios de validación en diferentes poblaciones, adaptaciones en español, en otros idiomas oficiales en España y versiones en el español hablado en diversos países de América, estudios de validación en población española y, ocasionalmente, valores de referencia en población general. En su mayor parte, se trata de instrumentos registrados cuya utilización requiere la autorización de sus propietarios (para usos en investigación esta autorización suele ser gratuita y ocasionalmente puede obtenerse a través internet).

Los **índices** son medidas derivadas de la teoría económica de la decisión. La calidad de vida se mide de modo global con un simple número que representa un valor en un continuo, generalmente de 0 (muerte o peor situación posible) a 1 (salud completa). Para su aplicación en estudios clínicos, existen dos aproximaciones fundamentales. En la primera se pregunta al paciente una serie de cuestiones sobre sus funciones físicas, sociales, etc. En base a sus respuestas, se clasifica al paciente en una determinada categoría que, a su vez, está asociada a un beneficio o utilidad, que ha sido previamente establecido en otros grupos. Este enfoque caracteriza a los instrumentos conocidos como escalas de bienestar y el más utilizado en nuestro entorno es, probablemente, el EuroQol (también dispone de adaptación española y estudios de validación). La segunda aproximación recoge en una simple ratio o en un valor numérico las preferencias del paciente respecto a diferentes estados de salud. Aquí los métodos más utilizados han sido la "lotería" estándar (**standard gamble**) y la preferencia temporal (**time trade-off**).

La mayor ventaja de los índices es su adecuación a los análisis coste-utilidad en la evaluación económica. De forma simple, se trata de ponderar el tiempo de vida ganado con cada intervención por la calidad de vida medida por el índice para construir **Años de Vida Ajustados por Calidad** (Quality Adjusted Life Year, QALY, AVAC). Por ejemplo, si un trasplante hepático permite

prolongar la vida de una persona concreta 10 años, los tres primeros meses con baja calidad de vida (supongamos un valor de 0,30), los 9 siguientes con una calidad intermedia (0,50) y el resto con una calidad aceptable (0,90), la agregación de estos valores $[(0,25 \text{ años} * 0,30) + (0,75 \text{ años} * 0,50) + (9 \text{ años} * 0,90)] = 8,55 \text{ AVAC}$ ofrecería el número de AVAC ganados.

En teoría, los AVAC pueden ser útiles para asignar recursos hacia aquellos programas de salud o intervenciones que mayor bienestar produzcan (medido en función de las preferencias de las personas encuestadas). Entre las limitaciones, las utilidades pueden variar en función de cómo fueron obtenidas, y se cuestiona la validez de una simple medida agregada. Además, y para su uso en clínica, no permiten determinar qué aspectos de la calidad de vida son responsables de los cambios y son menos sensibles que los perfiles para detectar pequeños cambios clínicos^{22,23}.

Otra alternativa en la medida de la calidad de vida es focalizar sobre aspectos que se consideren relevantes para un determinado problema de salud y aumentar la sensibilidad del instrumento sobre esa enfermedad en concreto. Este es el objetivo de los instrumentos específicos que pueden partir de una relación previamente establecida, o bien centrarse en los problemas que refiere el individuo.

La especificidad de un instrumento puede referirse a un tipo de enfermedad (dirigido a patologías oncológicas, cardíacas o respiratorias), pero también puede orientarse a una determinada población (como los diseñados para evaluar la calidad de vida de los ancianos que presentan una gran variedad de problemas), a una función en especial (cuestionarios que examinan la función emocional o sexual), o bien ser específicos de una condición o problema común para diferentes enfermedades (por ejemplo, el dolor). De todas estas aproximaciones, la que ha motivado un mayor desarrollo de instrumentos es la dirigida a medidas específicas de enfermedad que han proporcionado una considerable utilidad en estudios clínicos. Sus mayores

desventajas son la dificultad para utilizarlos en condiciones diferentes para las que fueron diseñados (dada la deliberada ausencia de diversos aspectos de la calidad de vida) por lo que generalmente se aplican junto a un instrumento genérico.

En la actualidad es posible encontrar centenares de instrumentos específicos. Entre los de mayor interés, con versiones validadas en español, cabe citar el Cuestionario de Calidad de Vida de la EORTC QLQ-30 (dirigido a pacientes de cáncer, que incluye un módulo general, para todos los cánceres, y módulos específicos para cáncer de mama, pulmón, esófago, cabeza y cuello, cuidados paliativos, colo-rectal y otros).

Es posible acceder a la mayoría de cuestionarios en español en [BiblioPRO](#), una biblioteca virtual de cuestionarios de CVRS y otros resultados percibidos por los pacientes (Patient Reported Outcomes, PRO) de acceso libre dirigida y desarrollada por el Grupo de investigación en Servicios Sanitarios del IMIM-Hospital del Mar, del CIBER en Epidemiología y Salud Pública (CIBERESP).

4. De la evaluación a la gestión: operativizando las medidas de resultado

Los indicadores de actividad o funcionamiento (performance indicators) son, en esencia, **criterios de evaluación aplicados a un conjunto de organizaciones o proveedores sanitarios**. Aunque cada conjunto particular de indicadores implica una concepción explícita de las expectativas sobre el funcionamiento del sistema sanitario o de algunos de sus componentes (hospitales u otros proveedores, áreas de atención, etc.), existe un amplio consenso en no separar las medidas de productividad, de calidad y de resultados clínicos, de modo que los indicadores de funcionamiento no se limitan a los de actividad y utilización de servicios, sino que incluyen una extensa gama de indicadores clínicos²⁴.

La utilidad básica de los indicadores de funcionamiento **es**

proveer información útil a los diferentes agentes del sistema sanitario para facilitar sus elecciones e intentar que sus expectativas se vean satisfechas. Dado que los diferentes agentes -y grupos sociales- tienen expectativas diferenciales sobre los servicios sanitarios, los indicadores deseables pueden variar ostensiblemente. Hay paquetes de indicadores que sólo son relevantes para algún componente del sistema (usualmente, y dada la naturaleza técnica de la atención sanitaria, instituciones sanitarias y profesionales).

No obstante, la tendencia actual es el desarrollo de paquetes estructurados de indicadores orientados -en algún grado- hacia los problemas de los pacientes y las comunidades, y que intentan ser tanto instrumentos esenciales de las políticas de mejora, como mecanismos para rendir cuentas a la sociedad de la eficiencia, efectividad y calidad con que se gestionan los servicios de atención de salud.

La utilización básica de los indicadores es comparar proveedores de servicios -u otros componentes del sistema- en un momento del tiempo o a lo largo del tiempo. En su forma más interesante, un indicador de funcionamiento no es más que la tasa de un determinado suceso -numerador- que forma parte de la atención prestada por los proveedores sanitarios a poblaciones definidas -denominador- en un tiempo definido.

Los sucesos pueden ser unidades monetarias, formar parte de los procesos de cuidados -incluyendo los servicios prestados- o tratarse de resultados (outcomes) de la atención médica. Los proveedores pueden ser médicos individuales o, más frecuentemente, centros sanitarios, hospitales, organizaciones sanitarias o territorios bajo una misma autoridad sanitaria. La población puede estar definida en función de su localización geográfica, aunque lo habitual es definirla también en función de alguna característica específica (diabéticos, hipertensos, etc.) o del riesgo respecto a un resultado de interés.

Las comparaciones transversales de proveedores utilizando indicadores se conocen como **perfiles de proveedores (profiling)**, término que define la aplicación de métodos epidemiológicos a la descripción de las prácticas asistenciales, la monitorización de resultados de la atención de salud y la evaluación de la calidad y la eficiencia de los cuidados, con los objetivos de proporcionar a gestores, compradores, usuarios y responsables de la toma de decisiones, información para comparar el coste, la utilización y la calidad de los proveedores, y proporcionar a los propios proveedores una referencia de sus costes y calidad respecto a otros proveedores, o respecto a su evolución temporal.

Por ejemplo, las tasas de administración de ácido acetil salicílico (suceso) por distintos equipos de atención primaria (proveedores), en pacientes remitidos al hospital por sospecha de infarto agudo de miocardio (población) en un año determinado (tiempo), definen el perfil de los proveedores de atención primaria respecto a un aspecto de la asistencia en el infarto agudo, y permite diseñar estrategias de mejora. Esta mecánica es aplicable a indicadores de resultados (tasa de complicaciones de herida en intervenciones de cirugía mayor programada, reintervenciones no planificadas, reacciones adversas a medicamentos, etc.), de productividad (estancia media ajustada por case-mix), de costes (por proceso, por habitante de un territorio), de satisfacción, de accesibilidad, etc.

El **benchmarking**, aunque requiere la realización de perfiles, incluye la **identificación de las mejores prácticas para dar una idea de los beneficios que se podrían alcanzar si las diversas organizaciones se situarán en el nivel de las mejores**. Aunque puede incorporar análisis más detallados para mostrar que es lo que hacen mejor otros centros y facilitar el aprendizaje, en el análisis de indicadores suele limitarse al establecimiento de un estándar para cada criterio definido por los centros que muestran los mejores resultados en ese criterio (por ejemplo, utilizar como norma deseable de duración de la

estancia en un proceso concreto, el promedio del 5 o 10% de los centros con mejor comportamiento, en lugar del promedio de todos los centros).

Referencias bibliográficas

1. Marión Buen J, Peiró S, Márquez Calderón S, Meneu de Guillerna R. Variaciones en la práctica médica: importancia, causas e implicaciones. *Med Clin (Barc)*. 1998;110(10):382-90.
2. Freund D, Lave J, Clancy C, Hawker G, Hasselblad V, Keller R, et al. Patient Outcomes Research Teams: contribution to outcomes and effectiveness research. *Annu Rev Public Health*. 1999;20:337-59.
3. Iezzoni LI. An introduction to risk adjustment. *Am J Med Qual*. 1996;11(1):S8-11.
4. Iezzoni LI. The risks of risk adjustment. *JAMA*. 1997;278(19):1600-7.
5. Meneu R. Top be not top be. *Rev Calidad Asistencial*. 2001;16:83-5.
6. Donabedian A. The role of outcomes in quality assessment and assurance. *Qual Rev Bull*. 1992;18:356-60.
7. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44:166-206.
8. Peiró S, Bernal-Delgado E. Hospitalizaciones evitables. ¿Quién soporta la carga de la prueba? *Rev Calidad Asistencial*. 2006;21:173-75.
9. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care*. 2000;12:281-95.
10. Fitzpatrick R, Fletcher A, Gore S, et al. Quality of life measures in health care. I: Applications and issues in assessment. *BMJ*. 1992;305:1074-77.
11. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118:622-9.
12. Eddy DM. Performance measurement: problems and solutions. *Health Aff (Millwood)*. 1998;17:7-25.

13. Kassirer JP. *The use and abuse of practice profiles*. *NEJM*. 1994; 330:634-6.
14. Hulka BS, Zyzanski SJ, Cassel JC, Thompson SJ. *Satisfaction with medical care in a low income population*. *J Chron Dis* 1971;24: 661-73.
15. Hulka BS, Zyzanski SJ. *Validation of a patient satisfaction scale: theory, methods and practice*. *Med Care*. 1982;20(6):649-53.
16. Ware JE Jr, Davies-Avery A, Stewart AL. *The measurement and meaning of patient satisfaction*. *Health Med Care Serv Rev*. 1978;1(1):1, 3-15.
17. Ware JE, Hays RD. *Methods for measuring patient satisfaction with specific medical encounters*. *Med Care* 1988; 26: 393-402.
18. Parasuman A, Zeilthaml V, Berry L. *A conceptual model of service quality and its implications for future research*. *J Marketing*. 1985;49(4):41-50.
19. Parasuman A, Zeilthaml V, Berry L. *Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria*. *J Marketing*. 1994;70(3):201-30.
20. Alonso J. *La medida de la calidad de vida relacionada con la salud en la investigación y la práctica clínica*. *Gac Sanit*. 2000; 14:163-7.
21. Badía X, Salamero M, Alonso J. *La medida de la salud. Guía de escalas de medición en español*. Barcelona: Tecnología y Ediciones del Conocimiento, 2007.
22. Guyatt GH, Naylor CD, Juniper E, et al. *Users' guides to the medical literature: XII. How to use articles about health-related quality of life*. *JAMA*. 1997;277:1232-7.

23. *Testa MA, Simonson DC. Assessment of quality of life outcomes. NEJM. 1996;334:835-40.*

24. *Peiró S. De la gestión de lo complementario a la gestión integral de la atención de salud: gestión de enfermedades e indicadores de actividad. En: Ortun V, ed. Gestión clínica y sanitaria: de la práctica a la academia, ida y vuelta. Barcelona: Masson; 2003:17-87.*