

Cuadernos de Evaluación Sanitaria



Número 5

Medidas de resultados en salud y su
incorporación a la evaluación económica

INSTITUTO MAX WEBER Y FUNDACIÓN GASPAR CASAL

© IMW & FGC

1st ed. NIPO

Cuadernos de Evaluación Sanitaria

Dirigidos por :

Álvaro Hidalgo, Director del Seminario de Investigación en Economía y Salud (SIES) Universidad Castilla la Mancha (UCLM)

Juan del Llano, Director de la Fundación Gaspar Casal (FGC) y Presidente de la Asociación Española de Evaluación de Tecnologías Sanitarias (AEETS)

Consejo asesor:

José Luis Poveda, Presidente Sociedad Española de Farmacia Hospitalaria (SEFH)

Miguel Ángel Calleja, Vocal Formación SEFH

Equipo editorial:

Juan Oliva. SIES UCLM

Isaac Aranda. SIES UCLM

Santiago Pérez Camarero. Director IMW

Flor Raigada. FGC

Gema Pi. FGC

Renata Villoro. IMW

Ángel Sanz Granda. Weber Economía y Salud, S.L

ÍNDICE

MEDIDAS DE RESULTADOS EN SALUD Y SU INCORPORACIÓN A LA EVALUACIÓN ECONÓMICA

1. Editorial. *Miguel Ángel Calleja*
2. Medidas de resultados en salud, clases y su aplicación en los estudios de evaluación económica. *Jordi Gol*
3. Medidas de salud reportadas por las pacientes: metodología, empleo y correlación con indicadores clínicos. *Montserrat Ferrer*
4. El AVAC como indicador de la esperanza y la calidad de vida. *José María Abellán*
5. Aspectos de interés para el SNS. *Isaac Aranda*
6. Hoja de síntesis práctica. *Isaac Aranda*



*Dr. Miguel Ángel Calleja
Hernández*

*Unidad de Gestión Clínica Provincial de Farmacia de
Granada Complejo Hospitalario Universitario de Granada.
Hospital Universitario Virgen de las Nieves*

EDITORIAL

La evaluación de las nuevas tecnologías desde los Servicios de Farmacia es una práctica muy frecuente por nuestro liderazgo en las Comisiones de Uso Racional del Medicamento y en la aplicación de criterios de uso adecuado del medicamento a la selección, evaluación y seguimiento del uso de los medicamentos.

Los nuevos medicamentos, cada vez más complejos desde el punto de vista tecnológico y considerados, por tanto, como nueva tecnología terapéutica aplicable en el entorno sanitario, son la herramienta terapéutica de mayor uso en nuestro sistema, generan innovación y mejor estado de salud en los pacientes, pero deben ser evaluados con una metodología rigurosa, científica y sistemática para garantizar la mayor seguridad para los pacientes, mucho más conociendo que suelen llevar consigo un coste mayor a medicamentos anteriores y, en algunos casos, podría comprometer el presupuesto del hospital, como es el caso de determinadas patologías oncológicas, reumatológicas o infecciosas, entre otras.

Los Servicios de Farmacia también estamos liderando y debemos seguir realizándolo, la implantación de nuevas tecnologías en los servicios que permiten realizar el trabajo de una forma más eficiente y seguro, como es el caso de la automatización de nuestras áreas logísticas con diferentes tipos de sistemas semiautomáticos y robotizados para la gestión de nuestros almacenes de medicamentos en la preparación de la medicación destinada a pacientes ingresados, pedidos de reposición de botiquines y en nuestras áreas de consultas externas. Esta área ha sufrido en los últimos años un cambio muy significativo y una mejora importante en la calidad de nuestro trabajo.

Además, nuestras áreas de preparación de medicamentos, están siendo mejoradas para la adaptación al RD 175/2001 con la incorporación de salas blancas y sistemas robotizados de preparación de medicamentos. Los Farmacéuticos de Hospital tenemos también que aplicar nuestra capacidad para realizar una adecuada selección del mejor medicamento a estas nuevas tecnologías que aplicamos en nuestros servicios de farmacia para garantizar la eficiencia también en el funcionamiento de nuestros servicios.

La incorporación de nuevas tecnologías en nuestras áreas clínicas y de seguimiento farmacoterapéutico ha revolucionado nuestra actividad diaria y ha mejorado nuestro seguimiento clínico de los pacientes al poder realizar un mayor número de tareas que garantizan la mayor seguridad para el paciente y la consecución de los mejores resultados en salud: el uso de la prescripción electrónica asistida y dotada de una información interrelacionada con bases de datos clínicas de parámetros de laboratorio y bases de datos de interacciones medicamentosas, dosis máxima y mínima, uso en situaciones especiales como insuficiencia renal, hepática, lactancia, embarazo, entre otras; ha supuesto un cambio en nuestra actividad y en la rentabilidad clínica de la misma.

Los Servicios de Farmacia también estamos liderando y debemos seguir realizándolo

De mismo modo las áreas de farmacocinética para la monitorización terapéutica de nuestros pacientes y para la predicción de respuesta a fármacos como la farmacogenética, también nos exige estar preparados para la evaluación y análisis económico de la idoneidad de la inversión en estos campos de desarrollo de nuestra profesión.

Una de las grandes áreas de mejora que tenemos en nuestro sistema es la evaluación de resultados en salud

Una de las grandes áreas de mejora que tenemos en nuestro sistema es la evaluación de resultados en salud. Con frecuencia adaptamos e implantamos nuevas tecnologías en nuestro entorno que no van asociadas a una evaluación de los resultados en salud que generan y que nos pueden ayudar a conocer la salud real que han generado en nuestros pacientes.

En este capítulo podremos conocer las diferentes herramientas que nos permiten cuantificar los resultados en salud. Conoceremos las diferentes herramientas disponibles y su adecuada aplicación, según cada caso, para conseguir los mejores resultados y poder aplicarlos a nuestras evaluaciones de medicamentos en particular y de nuevas tecnologías en general. Resultados en salud que podrán ser medidos por profesionales sanitarios ayudados de modernas tecnologías pero también podrán ser reportados por el propio paciente, siendo de ayuda inestimable para nuestra labor clínica directa con los pacientes, en especial para el seguimiento de patologías crónicas en las que el paciente, debidamente informado por los diferentes profesionales sanitarios en un trabajo multidisciplinar, puede convertirse en una paciente experto e informado, gran aliado para la consecución de los mejores resultados en salud.

Un paciente experto e informado, gran aliado para la consecución de los mejores resultados en salud

Además de la medición de la cantidad de las diferentes variables que es mejorada por el uso de los medicamentos, durante este capítulo, conoceremos también como medir la esperanza y calidad de vida con herramientas validadas y de calidad. Este tipo de sistemática tiene una especial relevancia, en el momento actual, para la evaluación de toda la innovación que están sufriendo las terapias en general y las biológicas en particular para el tratamiento de diferentes enfermedades, que suponen un impacto económico alto y en las que debemos asesorar a nuestras direcciones de los hospitales sobre la conveniencia de la inversión evaluando los resultados en salud así como su relevancia clínica (cantidad y calidad).

La aplicación de los conocimientos que se transmiten en este capítulo es directa en nuestra actividad diaria y su lectura y comprensión, sin duda, provocará una mejora importante en nuestro modo de proceder para realizar la evaluación de medicamentos, de nuevas tecnologías e incluso de nuestra forma de diseñar estudios clínicos de seguimiento y evaluación del impacto de la aplicación de nuevas tecnologías en nuestros servicios y en la terapéutica de nuestros centros hospitalarios.

Estos conocimientos provocarán una mejora en el farmacéutico de hospital en su vertiente asistencial, como agente que decide la implantación de una nueva tecnología en el hospital y a un paciente en particular; pero también la mejora se percibirá en las vertientes docentes e investigadora ya que nos permitirá un mayor desarrollo profesional en los campos de formación e información del uso de las nuevas tecnologías así como su evaluación previa a la implantación y posterior a la misma.

En definitiva, la aplicación de los conocimientos que se transmiten en este capítulo nos permitirá mejorar nuestras evaluaciones de medicamentos y de nuevas tecnologías y consolidarnos como los grandes aliados de nuestro sistema sanitario al garantizar la selección de las tecnologías más eficaces, seguras y eficientes para el conjunto del sistema sanitario en beneficio del paciente.



D. Jordi Gol-Freixa

Director de Programas en Instituto de Salud Carlos III

MEDIDAS DE RESULTADOS EN SALUD, CLASES Y SU APLICACIÓN EN LOS ESTUDIOS DE EVALUACIÓN ECONÓMICA

Resumen:

El trabajo realiza una descripción cronológica sobre el nacimiento de la medición de resultados en salud y su evolución a medida que la práctica clínica se desarrollaba. Así en el artículo se describe qué es necesario medir: “Death, Disease, Disability y Distress las cuatro D inglesas” y se recorre el camino que se ha seguido para poder cuantificar esos resultados desde el inicio hasta la actualidad.

1. Introducción desde la vida real

Varón de unos cincuenta y tantos años, casado y con hijos universitarios, con empleo estable de tipo técnico y ciertas responsabilidades directivas en una industria mediana. Acaba de pasar el reconocimiento médico anual que su empresa tiene contratada con cierta mutualidad con reputación de seria. Le han hecho una historia clínica, examen físico, análisis y una segunda visita de aconsejamiento. El hombre sale relativamente preocupado. Le ha subido el colesterol, que ya lo tenía alto; se confirma que tiene una hipertensión diastólica significativa. Ha aumentado algo de peso, “raspando la obesidad”. No ha comentado que ha tenido ciertos mareos ocasionales, ni detalles de sus preocupaciones familiares.

Lo que entendemos por salud –y también por enfermedad- ha ido evolucionando con el tiempo, con el conocimiento y con la tecnología y con la sociedad, incluidas las expectativas

Este ciudadano puede ser español, o de cualquier país del primer mundo. El tipo de herramientas que tienen los médicos a su disposición para ayudarle son comparativamente muy parecidas: las mismas moléculas o fármacos, equipos diagnósticos, revistas, libros de texto, e incluso criterios a la hora de actuar. La misma información en Internet. Pero, procede preguntarnos, si realmente va a recibir el mismo tratamiento en España, Canadá o Finlandia, es decir, que tipo de seguimiento se le va a hacer, por qué tipo de profesional, con qué medios. Hasta donde alcanza la modificación de estilos de vida, y por supuesto, cuál es el beneficio real, en términos de salud, de las distintas opciones que se le pueden plantear.

Lo descrito es bien familiar, es el pan nuestro de cada día en el primer mundo. Surgen muchas preguntas y debemos responderlas en orden. La primera que deberíamos plantearnos es: el ciudadano, ¿está realmente enfermo?. Lo que entendemos por salud –y también por enfermedad- ha ido evolucionando con

el tiempo, con el conocimiento y con la tecnología y con la sociedad, incluidas las expectativas (1). En nuestro ejemplo, lo descrito son factores de riesgo, básicamente intrínsecos, pero no son una enfermedad en si misma. Ahora bien, hay que señalar que científicamente tienen cierta solidez cómo predictores de estados futuros de salud.

CUADRO 1.- NIVELES DE INTERVENCIÓN DE LEAVELL & CLARK, MODIFICADO (2).

PREVENCIÓN PRIMARIA	PREVENCIÓN SECUNDARIA	PREVENCIÓN TERCIARIA
Favorecer el estilo de vida saludable	Diagnostico precoz	Prevención de la iatrogénica,
Educación sanitaria	Intervención diligente	Rehabilitación
Prevención específica, social, étnica u ocupacional	Intervención de intensidad adecuada	Reinserción social

Entre 1959 y 1965 estos autores definieron y refinaron el concepto de prevención y lo incardinaron en el marco más amplio que incluye tanto la salud pública como la asistencia sanitaria. De aquella época son los primeros estudios de sociología de la salud modernos y el interés inicial por la sanidad como terreno de estudio por parte de las ciencias sociales. Evidentemente, este esquema conceptual parte desde la perspectiva de lo que tradicionalmente se conocía como “la historia natural de la enfermedad”, paradigma prevalente en los últimos 200 años. Se han introducido muchas críticas y matizaciones en los últimos veinte años, notablemente desde la sociología y desde la gestión clínica

Sabemos que los factores de riesgo se pueden medir desde los años setenta, década en que se empezó a generalizar el concepto de riesgo relativo, y muy especialmente a partir de la gran cohorte de Franingham, en la que se llevó a cabo el seguimiento sistematizado durante muchos años de toda una población mediante procedimientos normalizados, lo que permitió desarrollar el concepto de interacción de riesgos. Es en este punto donde se da el salto de la epidemiología

Los factores de riesgo se pueden medir desde los años setenta, el seguimiento sistematizado durante muchos años de toda una población mediante procedimientos normalizados, lo que permitió desarrollar el concepto de interacción de riesgos. Es en este punto donde se da el salto de la epidemiología descriptiva a la analítica.

descriptiva a la analítica. Este esquema conceptual ha estado vigente en los últimos cuarenta años, con diversos matices y variaciones.

En el caso de nuestro hombre se plantearán medidas de tipo farmacológico, como uso de hipolipemiantes y antihipertensivos. También de estilo de vida, que incluirán modificación de dieta y sedentarismo. Asimismo se considerará la necesidad de exploraciones adicionales. Es en este punto donde van a existir notables diferencias entre distintos contextos sociales y sanitarios. En algunos de ellos existirán programas integrales para el tratamiento y seguimiento de grupos de ciudadanos en condiciones parecidas, tal vez organizados por la propia empresa u organización patronal o mutualidad. En otros dependerá del criterio del médico de cabecera y los medios que tenga a su disposición. En un tercer caso nuestro hombre se encontrará en un mercado libre con multitud de medios a escoger, sujeto a una "oferta" tal vez agresivamente persuasiva. En algunos casos extremos se ha considerado que puede incluso ser "enfermiza" la oferta sanitaria cuando hay una base de afán de lucro, sea por parte de la industria (disease mongering(3)) o de los profesionales del sector, que de alguna manera expropian la experiencia vital. Son cuestiones filosóficas y políticas de calado, a no ignorar, pero que trascienden el marco de este artículo. Señalemos algún aspecto de este caso que resulta clave para plantear honestamente la cuestión de la medición de resultados. En primer lugar cuál es el resultado en sí, si es corregir un factor de riesgo el propio seguimiento clínico lo indicará. Si es evitar una enfermedad nos encontramos en que los plazos son importantes. Pero el punto principal es poner límites a la simplificación. En el estado del conocimiento actual sabemos que intervienen significativamente los factores mencionados, y que además hay interacción entre ellos, por este motivo no hay que simplificar más de la cuenta.

Los cuidados formales son aquellos prestados de manera profesional a personas con limitaciones en su autonomía con el fin de favorecer el desempeño de sus actividades básicas e instrumentales en su vida diaria.

2. “Back to Basics”

El beneficio que aportan los medicamentos está bien medido mediante series clínicas robustas, y, en general hay datos de las virtudes de todas las posibles intervenciones que pudieran beneficiar a nuestro hombre, aunque la calidad de esta información va a ser variable: es mucho más robusta la relativa a intervenciones farmacológicas que a los aspectos de modificación de estilos de vida; puede que parte de esta información no siempre sea de fácil acceso, especialmente cuando hay factores de tipo local relevantes. Más compleja es la tarea de cómo combinar esta base de evidencias, o el valor adicional de exploraciones o diagnósticos adicionales. Aun así, existe un cuerpo acumulado de revisiones sistemáticas de la evidencia, que en un porcentaje alto de patologías son de relevancia directa para el clínico, y en muchos casos además existen guías de práctica clínica validadas. Pero antes de entrar en las complejidades necesarias procedamos a establecer conceptos muy básicos, lo que los anglosajones llamarían “the bottom line”. Para ello, una regla mnemotécnica ha resultado útil, en inglés es la de las cuatro des (“D”). Tanto si hablamos de salud o de enfermedades, que son conceptos asimétricos, o de personas, o en tanto a miembros pertenecientes a poblaciones, podemos establecer un primer nivel de consenso sobre que los parámetros que interesan se pueden enmarcar en combinaciones de estos cuatro:

Existe un cuerpo acumulado de revisiones sistemáticas de la evidencia, que en un porcentaje alto de patologías son de relevancia directa para el clínico, y en muchos casos además existen guías de práctica clínica validadas.

1. “Death” mortalidad
2. “Disease” enfermedad, prevalencia, secuelas
3. “Disability” incapacitación, pérdida de autonomía
4. “Distress” sufrimiento

Evidentemente son conceptos o ideas muy básicos, y, precisamente por ello resultan útiles. Se podría intentar rastrear su origen hasta la antigüedad. Empezaron a tomar notoriedad en los orígenes de la salud pública organizada, conocida como “higiene” o sanidad en las épocas coloniales, hacia mitades del siglo XIX, fechas por las que se producen dos hechos de forma simultánea. En primer lugar, el saneamiento se convierte en una cuestión de Estado, en parte por motivos coloniales, en este periodo la potabilidad de las aguas y la salubridad son prioridades de los

Estos parámetros propuestos, “Death, Disease, Disability y Distress las cuatro D inglesas” deben, precisamente por lo básicos, considerarse siempre ante la evaluación de cualquier intervención sanitaria, del tipo que sea.

incipientes estados industriales. En segundo término, empieza a desarrollarse la estadística descriptiva sanitaria, donde Florence Nightingale fue a la vez precursora y activista.

Estos parámetros propuestos, “Death, Disease, Disability y Distress las cuatro D inglesas” deben, precisamente por lo básicos, considerarse siempre ante la evaluación de cualquier intervención sanitaria, del tipo que sea (4).

3. Tipología y validez de las medidas de resultados en salud.

Obviamente la primera es la propia mortalidad, y la que ha merecido mayor atención en los últimos dos siglos, tanto desde la demografía como los primeros estudios de higiene. Es el resultado final indiscutible, lo que en lenguaje de investigación se llama el “endpoint”. El término “end-point” es importante, y lo es por motivos de tipo conceptual, de tipo metodológico, de tipo

La tasa de mortalidad intrahospitalaria, fue considerado el indicador más sensible de gravedad de las patologías, antes de la introducción de la clasificación de procedimientos que supuso la implantación de los GRD

práctico y de tipo ético. Cuando comparamos, por ejemplo, distintos programas de cirugía cardíaca tenemos multitud de parámetros a evaluar, pero no se debe obviar a priori la tasa de mortalidad global. Tradicionalmente la tasa de mortalidad intrahospitalaria, asumiendo igualdad de calidad en los servicios – que es mucho asumir- fue considerado el indicador más sensible de gravedad de las patologías, antes de la introducción de la clasificación de procedimientos que supuso la implantación de los GRD (Grupos Relacionados por el Diagnóstico, hacia 1982).

Una primera clasificación nos la da lo objetivo que pueda ser el parámetro, siendo la mortalidad el más obvio. Los datos de laboratorio y muchas pruebas diagnósticas se pueden considerar también objetivas. Muchas tienen algún componente de interpretación, por ejemplo, las imágenes radiológicas. Sin embargo, esto no ha sido siempre así. Ha existido un proceso continuado y sostenido de estandarización que ha redundado en una mejora progresiva de la consistencia. Hace veinte años existía gran variación entre los laboratorios, por la diversidad de técnicas y reactivos. Por ejemplo, los test de protrombina para indicar el estado de la coagulación. Este era el caso de muchas pruebas cuando los reactivos eran de base biológica. La aparición de técnicas recombinantes ha cambiado radicalmente esta situación, además de todos los procesos de normalización. El progreso de la informática aplicada, con técnicas de inteligencia artificial está contribuyendo a la automatización de muchos procesos previamente manuales, como recuentos celulares, citologías, e incluso, la cuantificación de imágenes digitales, que es relevante tanto en radiología como en anatomía patológica. Por todo ello, en la actualidad existen muchos más parámetros objetivos.

En este sentido siempre hay que considerar qué elementos de juicio profesional o científico hay que realizar. El juicio clínico ha sido motivo de reflexión desde hace mucho tiempo. Tomó especial novedad con la generalización de los métodos diagnósticos complementarios –laboratorio y radiología, sobre todo- junto con el aumento de las interconsultas entre especialidades; Marañón, Laín trataron la cuestión con profundidad considerable: siempre es buena práctica

Un problema práctico en la investigación eran las variaciones entre los observadores, para lo cual se desarrollaron métodos estadísticos y procedimientos para tratar la concordancia entre los observadores independientes.

la existencia de una segunda opinión, siendo ésta un privilegio del paciente y su entorno ante decisiones trascendentes, y mandataria en los proyectos de investigación. Y es importante que esta segunda opinión sea independiente, no “contaminada”. Con la aplicación de la epidemiología a la clínica, a partir de 1975-80 hubo un desarrollo considerable en la formalización del juicio clínico (5). Un ejemplo fueron los “criterios de Jones” para el diagnóstico de la fiebre reumática, revisados en base cuantitativa hacia 1980, que coincidió con la sistematización de los síndromes, el

desarrollo de la historia clínica organizada por problemas (“problem oriented medical record”), la protocolización de la interpretación de las imágenes radiológicas, y los inicios del “medical decision making”. Un problema práctico en la investigación eran las variaciones entre los observadores, para lo cual se desarrollaron métodos estadísticos y procedimientos para tratar la concordancia entre los observadores independientes.

3.1. Epidemiología y validez.

La generación de los datos corresponde a la epidemiología, término que combina las tres raíces griegas: epi, por sobre, demos, población, logos, entendimiento. Es una ciencia aplicada, y al ser aplicada también un arte, que va más allá y está encima de la bioestadística, disciplina que para la epidemiología es necesaria e instrumental. Se consideran especialmente tres tipos de cuestiones de validez, desde la epidemiología. Son la validez externa, que hace referencia a hasta donde se pueden generalizar los datos de un determinado

La epidemiología es una ciencia aplicada, y al ser aplicada también un arte, que va más allá y está encima de la bioestadística, disciplina que para la epidemiología es necesaria e instrumental

estudio. La validez interna, que hace referencia al rigor y la calidad del estudio. En tercer lugar, la validez de constructo, que es especialmente importante cuando las variables independientes son psicosociales, y que tiene que ver con la consistencia y mensurabilidad de los conceptos en estudio.

Validez interna.
El problema principal es la identificación y en su caso, el control de los sesgos

3.1.1 Validez interna.

El problema principal es la identificación y en su caso, el control de los sesgos. Es decir, que por el diseño o ejecución del estudio haya algún tipo de error más o menos sistemático. En una revisión histórica realizada hacia 1975 se llegaron a identificar hasta más de un centenar de posibles causas de sesgo descritas en la literatura de la época. No todos los sesgos tienen el mismo alcance, en primer lugar estarían los problemas con los controles, como por selección inadecuada, por riesgo, patología, geografía. En segundo término, el sesgo de Bergson, que consiste en que si los controles son hospitalarios puede darse una correlación que tenga que ver con la patología prevalente en el área y el centro actuar como seleccionador de riesgo. En tercer lugar, uno de los más importantes, el

El esquema más robusto es cuando se selecciona una población bien definida y se sigue sistemáticamente a lo largo del tiempo, antes de que se manifieste la enfermedad clínica, lo que se denomina una cohorte poblacional, que permite además establecer datos fiables de incidencia.

sesgo de rememorización ("recall bias"): los pacientes o casos recordarán mucho mejor la exposición a un factor de riesgo que los controles. Para identificar los sesgos potenciales, el punto clave es la direccionalidad temporal del estudio: si éste es prospectivo o retrospectivo. Un estudio retrospectivo consiste en identificar unos casos de patología y aparearlos con sujetos "controles", comparando la exposición a factores de riesgo. O sea, la información acerca de la exposición hay que reconstruirla a posteriori, con los muchos y diversos

problemas que conlleva. En un estudio prospectivo se define “a priori” la información a recoger y se desarrolla una sistemática para su captura y almacenamiento, tanto para casos como controles. El esquema más robusto es cuando se selecciona una población bien definida y se sigue sistemáticamente a lo largo del tiempo, antes de que se manifieste la enfermedad clínica, lo que se denomina una cohorte poblacional, que permite además establecer datos fiables de incidencia. Un ejemplo de ello fue el estudio de los factores de riesgo en Framingham, Massachussets, que significaron un cambio de paradigma en el entendimiento de los factores de riesgo cardiovasculares. O el estudio de Karelia del Norte, Finlandia, que permitió modelar distintos aspectos de la nutrición y estilos de vida. Son estudios realmente caros.

En el caso de las intervenciones sanitarias está el efecto de los propios profesionales y el efecto indirecto del sistema sanitario, lo que se ha venido a llamar “efecto placebo”, que se superponen a los sesgos de selección de pacientes. Para ello se ha desarrollado un diseño especial, el llamado “ensayo clínico” en el que se compara el efecto de una intervención, en general un medicamento, con algún comparador. Se puede considerar el ensayo clínico como un caso especial de estudio prospectivo. A efectos de controlar la

Hay dos desarrollos que es importante mencionar. El primero es la combinación y síntesis de resultados de los estudios experimentales, que se ha venido a denominar meta-análisis, que es una técnica estadística y analítica para combinar los resultados de los distintos estudios, ponderándolos por su potencia estadística. Un segundo aspecto es el de las revisiones sistemáticas de la literatura

selección de pacientes y riesgos, se identifican bien las condiciones de inclusión y de exclusión de sujetos para el estudio, que deben ser idénticas para ambos grupos, el experimental y el de control. Para controlar el efecto placebo, incluso el indirecto, se considera que el diseño es más robusto cuando hay un enmascaramiento del tratamiento entre pacientes y médicos, “blinding”, en inglés, ya que a efectos de la intervención, los clínicos no saben que sustancia están dando, lo cual implica mecanismos de control muy rigurosos; y para evitar sesgos de selección, es decir,

que hayan diferenciales de riesgo a priori, se hace una asignación al azar de los pacientes entre los dos grupos (“randomización”). Este tipo de estudios tiene sus problemas éticos propios, en parte dependientes de la incertidumbre, lo que obliga a supervisión y regulación específica, con revisión previa de protocolos y tutela por comités de ética. Un aspecto que no siempre se resalta es que ambas opciones, el tratamiento experimental y el de comparación deben ser éticamente aceptables “a priori”, con lo cual no es prudente esperar resultados muy espectaculares.

En las últimas décadas, en términos de volumen de evidencia o de datos de estudios se estima que son diez veces más los provenientes de ensayos clínicos que de cohortes u otros estudios epidemiológicos no-experimentales. Otro aspecto adicional importante es el sesgo de publicación. Los estudios prospectivos suelen estar patrocinados por instituciones públicas o no lucrativas y se suelen publicar; la mayoría de ensayos clínicos con tecnologías tienen un promotor con intereses en el sector, que en parte está motivado por el posible lanzamiento de un nuevo fármaco o tecnología y que tiene el privilegio de publicar o no los resultados. Esto conlleva que la mayoría de los resultados negativos, es decir, que no muestran ventajas o incluso mostrarían desventajas, no se publican. Este sesgo de publicación queda recogido en el logotipo de la colaboración Cochrane. Cuando se han comparado los resultados de ensayos clínicos con estudios de cohortes bien diseñados y ejecutados se ha visto que las diferencias no son significativas. Ante la proliferación de estudios prospectivos hay dos desarrollos que es importante mencionar. El primero es la combinación y síntesis de resultados de los estudios experimentales, que se ha venido a denominar meta-análisis, que es una técnica estadística y analítica para combinar los resultados de los distintos estudios, ponderándolos por su potencia estadística. Un segundo aspecto es el de las revisiones sistemáticas de la literatura. Ambos se han ido transformado progresivamente en herramientas para la toma de decisiones en política sanitaria y de gestión, y en general, de un indicador del estado del conocimiento basado en la evidencia. Una

Las revisiones sistemáticas son una de las tareas centrales para la evaluación de tecnologías en salud (ETS), y para el desarrollo de las guías de práctica clínica (GPC)

revisión sistemática implica, como su nombre indica, una compilación de toda la información existente de fuente experimental que cumpla unos criterios de inclusión (relevancia) y exclusión (falta de rigor). Las revisiones sistemáticas son una de las tareas centrales para la evaluación de tecnologías en salud (ETS), y para el desarrollo de las guías de práctica clínica (GPC). Todo esto ha implicado un esfuerzo importante, en parte motivado por los propios intereses genuinos de los propios clínicos investigadores, y por otra parte, de las administraciones públicas, y en algún caso alguna aseguradora privada, que han visto deseable disponer de capacidad de análisis y juicio propio, creando para tal fin organismos autónomos para el apoyo a la política sanitaria, priorización de inversiones y revisión de cobertura de servicios: las agencias de evaluación de tecnologías. La colaboración interprofesional, paralela al movimiento surgido en los años 1990 de la medicina basada en la evidencia ha conducido al desarrollo de la iniciativa de la Cochrane Library, que compila de forma sistemática una biblioteca de ensayos clínicos publicados, habiendo organizaciones paralelas como repositorios de revisiones sistemáticas, y de guías de práctica clínica.

Para ciertos tipos de patología o intervenciones que tienen ciertas peculiaridades de incidencia, no es planteable la randomización ni es práctica una cohorte de incidencias de tipo poblacional, como es el caso de los trasplantes o del cáncer. El abordaje es el registro de casos incidentes

La cuestión de la validez interna también se presenta en los estudios de prevalencia, que básicamente son las encuestas de salud. Aquí los métodos de recogida y tratamiento de datos, estratificación de sub-muestras de población y otros aspectos metodológicos prácticos se han beneficiado lógicamente de los desarrollos realizados en los estudios experimentales. Recordemos que los estudios de prevalencia son caros, y que su diseño, para que sean robustos, debe ser bastante sofisticado, ya que combinan problemas de validez interna con validez externa. Un sesgo específico es que existe siempre una correlación entre los pacientes que no se encuentran, o no responden siendo los de situación social y sanitaria más precaria.

Para ciertos tipos de patología o intervenciones que tienen ciertas peculiaridades de incidencia, no es planteable la randomización ni es práctica una cohorte de incidencias de tipo poblacional, como es el caso de los trasplantes o del cáncer. El abordaje es el registro de casos incidentes, y si el diseño es bueno y la gestión excelente las coberturas pueden ser notables, llegando casi al 100% en enfermos renales o todo tipo de trasplantes, o sobrepasando el 80% en los casos incidentes de cáncer. Es la única estrategia posible para el estudio de los factores de riesgo de las enfermedades de baja prevalencia, a veces llamadas “enfermedades raras”.

Se suele utilizar una gradación de tipos de estudios según robustez, que puede tener algún valor didáctico, pero en la que conviene no olvidar una serie de matices. De menos a más robusto, la escala sería la del cuadro 2. Es a todas luces una simplificación necesaria, y en general la mayoría de instituciones tienen la suya.

CUADRO 2.- ROBUSTEZ DE DISEÑOS EN ORDEN CRECIENTE.

1	Series clínicas cortas de un centro, definidas a posteriori
2	Series clínicas bien definidas, con controles retrospectivos
3	Estudios prospectivos de casos y controles.
4	Cohortes poblacionales donde se pueda definir la incidencia
5	Ensayos clínicos controlados (criterios de inclusión y exclusión “a priori”)
6	Ensayos clínicos randomizados, grandes.
7	Metaanálisis/revisión sistemática de varios ensayos clínicos

Fuente: elaboración propia. Esta escala marca la “calidad de la evidencia” de eficacia o del estado del conocimiento.

Durante mucho tiempo se ha argumentado que a la hora de emitir una recomendación esta debería ser proporcional a la calidad de la evidencia, habiendo existido un considerable debate desde 1990 hasta hace pocos años. El origen en parte estaba alimentado por la renovación que significó en su tiempo la aparición del movimiento “medicina basada en la evidencia” (EMB) y fue originariamente propuesto por una organización nueva, un grupo de trabajo de universidades americanas punteras que crearon el **U.S. Preventive**

Services Task Force (USPSTF) (6). El relacionar la recomendación a la calidad de la evidencia tiene su lógica, que es la progresiva reducción de incertidumbre al aumentar el volumen y masa de evidencias, claro, pero es importante no olvidar el dicho de “la falta de evidencias de eficacia no implica evidencia de ineficacia”. La penicilina G o las sulfamidas o los anticoagulantes fueron introducidos en el mercado –otras épocas, ciertamente- sin apenas investigación clínica en humanos. A medida que la introducción de los medicamentos es más actual, lógicamente hay más estudios y más datos –evidencias- conforme los medicamentos o las tecnologías son más modernas, ya que así queda incentivado por el propio sistema regulatorio.

La investigación clínica en medios diagnósticos es un caso especial, donde se plantean muchas de las cuestiones de validez interna, pero también de constructo y de validez externa. No suele ser viable el enmascaramiento, lo cual se palia en parte mediante la rigurosa independencia en la evaluación de las distintas pruebas en comparación. No hay a priori necesidad de grupos control, todos los pacientes pueden recibir ambas exploraciones, hechas de forma completamente ciega a los resultados de la otra.

En el caso de los procedimientos quirúrgicos, a pesar de alguna iniciativa anecdótica tampoco se puede plantear el enmascaramiento, y la cuestión de la asignación aleatoria tiene más variables, ya que existe una relación calidad/volumen marcada. Por ejemplo, cuando se ha intentado comparar la cirugía coronaria con la angioplastia, de 1000 pacientes potenciales, una vez aplicados los criterios de exclusión de ambas técnicas quedaban como randomizables menos de una quinta parte, siendo los resultados en estos casos muy comparables, lo cual no es de sorprender. En el caso de los procedimientos quirúrgicos, el disponer de un buen registro y excelente monitorización, como el de la Soviet of Thoracic Surgeons, ha resultado ser muy eficaz para la mejora sostenida de resultados.

La cuestión de la robustez de los métodos de evaluación de la validez interna, tal como se ha mencionado, ha dado mucho que discutir, dadas las implicaciones regulatorias, y por supuesto comerciales de los promotores, y “last but not least”, la propia integridad de los investigadores. Se han propuesto distintos tipos de criterios de medición de la calidad de los mismos, con escalas más o menos prolijas. Una revisión de 1999 tomó en consideración más de una docena de propuestas, alguna incluía hasta 22 ítems, no exentas de inconvenientes (7). Un trabajo muy ingenioso es el de A. Jaddad que revisando la calidad de los ensayos clínicos sobre el dolor vio que los criterios se podían reducir a 5 (8).

La pregunta ante cualquier estudio es la descomponían en dos partes: son los resultados válidos, y son de aplicación a mi paciente.

3.1.2 Validez externa. Población de estudio y población diana.

En este apartado vamos a comentar hasta dónde se pueden generalizar los resultados obtenidos. Un ejemplo: media docena de vacunas para la prevención de la gripe fueron evaluadas en su fase III en la República Sudafricana, siendo la población de una edad media, de profesión minero, de raza Zulú o Xhosa y grupo sanguíneo mayormente B, 90% varones. ¿Es legítimo extrapolar estos resultados a menores de edad escandinavos? ¿O a jubilados españoles? En este caso, claramente no. David Sackett y Gordon Guyatt, principales promotores de la Medicina Basada en la Evidencia (MBE) en su día generaron un criterio especialmente elegante para la evaluación de intervenciones terapéuticas. La pregunta ante cualquier estudio es la descomponían en dos partes: son los resultados válidos, y son de aplicación a mi paciente. El criterio era así de simple: mi paciente en concreto hubiera sido elegible para participar en este estudio. Si la respuesta es afirmativa, es válido, sino, no.

En este sentido y a la hora de generalizar los resultados, hay recordar que los resultados no son uniformes, ni en parámetros de laboratorio ni clínicos, que hay algún tipo de distribución, muy frecuentemente de tipo curva normal modificada, y que se dan las medias e intervalos de confianza, con lo cual la

predicción de un resultado en un caso concreto siempre debe tomarse con reservas.

3.1.3 Validez de constructo.

Este es especialmente relevante cuando se trata de aspectos psicológicos o sociológicos, y tiene que ver con la consistencia del concepto en estudio y su forma de medirlo, normalmente una escala. Es de aplicación, por ejemplo, a todo lo que es experiencia subjetiva, como dolor, energía, autonomía, y, en general, en las escalas de calidad de vida u otras herramientas o instrumentos utilizados en los estudios coste-utilidad. Hay métodos desarrollados para evaluar su consistencia interna, robustez y reproducibilidad.

3.2 Calidad de las medidas.

Ya apuntábamos a las distintas escalas para medir la calidad de los ensayos clínicos y su elegante reducción al check-list de Alejandro Jaddad y los motivos para ello. En parte han estado también influidos por las necesidades de rigor de las propias revistas científicas: por ejemplo la primera en imponer criterios muy estrictos para la publicación de ensayos clínicos fue "The Lancet" que exigió la presentación de un diagrama de flujo explícito sobre la pérdida de follow-up de pacientes (que siempre es sospechoso de que algo no va bien). Posteriormente se han ido lanzando distintas iniciativas para marcar unos criterios de calidad metodológica, como indicamos en el cuadro 3º. Estas iniciativas, suelen ser promovidas por círculos académicos vinculados con el mundo de la literatura. Con diferencia, los más importantes son los terapéuticos.

A. Ensayos clínicos:

CONSORT STATEMENT (Consolidated Standards of Reporting Trials) (9).

B. Estudios observacionales:

STROBE (STrengthening the Reporting of OBServational studies in

Epidemiology) Statement (10).

C. Estudios diagnósticos:

Standards for the Reporting of Diagnostic Accuracy Studies (STARD) (11).

D. Metaanálisis y revisiones sistemáticas:

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (12).

E. Estudios para la asociación de patologías con la genética (desarrollo de los estudios observacionales)

Strengthening the Reporting of Genetic Association studies (STREGA) (13).

Como se puede apreciar, estas iniciativas están vinculadas a distintas redes promotoras de la calidad y la excelencia tanto en la investigación como en la publicación (“reporting”) (14). También cabe contemplar todas estas loables iniciativas desde el punto de vista del economista. Han contribuido a hacer mucho más rigurosa la investigación aplicada, en un momento donde las fuerzas de la oferta asistencial son más potentes que los propios reguladores, es decir los estados, tal como hemos comentado al principio, pero en una lógica comercial, más concretamente de márketing, no conviene olvidar que son barreras de entrada a la competencia, fidelización de colectivos y sectores, y en definitiva, el desarrollo de una industria auxiliar o proveedora de cierta entidad, los “evidenciólogos”, fenómeno cuya deriva ya se había apuntado hacia 1995 (15).

Por una parte se sigue confundiendo la potencia estadística con la magnitud del efecto, más de treinta años después de que esta falacia fuera descrita. Sigue existiendo la tendencia en presentar los resultados en cocientes de riesgos relativos (“relative risk ratios”, RRR) en vez de riesgos atribuibles

3.3 Potencia de las medidas.

Antes hemos apuntado a la tentación de vincular la fuerza o potencia de la recomendación a la calidad o robustez de la evidencia. En esta tentación han participado muchas instituciones, y ha sido tenida en cuenta sobre todo en la introducción de nuevas tecnologías, lo que algunos han llamado la cuarta barrera ("the fourth hurdle"). Ahora conviene comentar dos dificultades adicionales relevantes en la práctica del mundo real. Por una parte se sigue confundiendo la potencia estadística con la magnitud del efecto, más de treinta años después de que esta falacia fuera descrita. Sigue existiendo la tendencia en presentar los resultados en cocientes de riesgos relativos ("relative risk ratios", RRR) en vez de riesgos atribuibles. Un ejemplo cruel y brutal: aumentar la supervivencia de un cáncer metastásico a seis meses de 4% de la serie de pacientes a un 6% no es una mejora de un 50% tal como desgraciadamente se sigue planteando alguna vez. Es una mejora de un 2%, valiosa, sin lugar a dudas, pero en otra dimensión.

El tercer aspecto a mencionar es que estamos siempre tratando de datos obtenidos en situaciones asistenciales en las que ha existido a) un nivel alto de supervisión y monitoreo, con sus costes; b) un personal asistencial altamente entrenado, informado y motivado, en centros de excelencia, y c) una selección rigurosa de los pacientes. Por tanto tenemos una sobreestimación de los beneficios potenciales que ofrecemos a la población y a los profesionales, que en parte está exagerada por la propia dinámica de comunicación. El farmacoeconomista, el financiador, los profesionales, y el público en general de alguna manera lo saben, aunque no hay demasiados datos que documenten esta sobreestimación.

4. Limitaciones del abordaje académico. Eficacia frente a efectividad.

En muchas patologías graves la asistencia se realiza con niveles de supervisión y con protocolos de calidad encomiable. La cirugía con circulación extracorpórea, por ejemplo, o las intervenciones en unidades de cuidados

intensivos exigen una supervisión muy rigurosa que viene impuesta por la propia dinámica de la asistencia. En tales unidades asistenciales, los resultados reales en el día a día podrán ser parecidos a los resultados de los ensayos clínicos, o de los propios registros de calidad. Pero este no es el caso en muchos otros “settings”, a pesar de las considerables mejoras de las tecnologías de la información (TI) y de la documentación clínica. Además de los factores descritos antes, como los medios complementarios disponibles, la intensidad de la supervisión o la motivación intrínseca del personal hay otro tipo de factores. Por ejemplo, es difícil evaluar con cierta consistencia, la dermatología en atención primaria, o el tratamiento de patologías erráticas como puedan ser las úlceras crónicas, sean diabéticas o vasculares. En este sentido, hay que señalar que la información tiene un coste, en tiempo y en calidad.

Durante décadas se ha considerado el lugar común de que un 80% de las prácticas reales, en términos de volumen, y en términos económicos no tenían su eficacia medida. En parte ha sido esta sospecha la que alentó la creación de las primeras agencias de evaluación de tecnologías sanitarias.

Pero existe un factor adicional que es el cumplimiento terapéutico (“la compliance”), que conforme nos alejamos de la situación del ensayo clínico empeora. Baja con la polimedicación, con el aumento de edad, y ciertamente con las situaciones de integración social: pobreza, bajo nivel cultural, desestructuración familiar, incultura, desarraigo. La influencia de todos estos factores, en el cumplimiento terapéutico, está altamente descrita en la literatura. Por este motivo, tenemos que distinguir entre eficacia y efectividad. A este respecto, tenemos dos fuentes de información, la académica, rigurosa, descrita, que nos aporta la información de eficacia. Y luego tenemos la información administrativa que nos aporta información de utilización o efectividad. Ambas han ido en paralelo hasta hace relativamente poco.

4.1 Cochrane y la cuestión de la efectividad.

Esta diferencia entre lo que se sabía ya de la eficacia en circunstancias ideales, por utilizar el lenguaje de la época, y de la efectividad en condiciones habituales ya fue plantada por Sir Archie Cochrane, por aquel entonces presidente del Medical Research Council del Reino Unido. Cochrane avanzaba su sospecha, razonablemente fundada de que buena parte de las prácticas médicas de la época tenían una efectividad dudosa y ciertamente no validada, introduciendo el término de effectiveness. Durante décadas se ha considerado el lugar común de que un 80% de las prácticas reales, en términos de volumen, y en términos económicos no tenían su eficacia medida. En parte ha sido esta sospecha la que alentó la creación de las primeras agencias de evaluación de tecnologías sanitarias.

Procede aquí apuntar una aclaración terminológica que en el pasado ha creado malos entendidos públicos. Y es la del distinto uso del lenguaje anglosajón médico, trasladado y traducido al español por la vía directa, frente al lenguaje jurídico y filosófico de amplia tradición en nuestro medio. En español los juristas usan el término *eficacia efectiva* cuando implican un nexo causal físico que es condición necesaria y suficiente. En el ámbito sanitario se usa el término *efficacy* traducido por *eficacia* cuando se hace referencia al efecto en circunstancias controladas y experimentales, y se usa el término *effectiveness*, traducido por *efectividad*, cuando se hace referencia al impacto en el mundo real. Obsérvese el matiz adicional de que la eficacia tiene una dimensión individual, y la efectividad, conceptualmente, es poblacional.

Hay además una cuestión moral que en tiempos de mejor bonanza económica ha sido obviada. No es ético financiar a expensas públicas medidas, o dispendios sanitarios cuya efectividad sea dudosa. Esto se empezó a plantear por primera vez a raíz de la crisis económica de los años noventas, sobre todo en los EEUU, cuando se tomó conciencia de que la información sobre actividad asistencial era completamente inadecuada con la excepción de alguna mutualidad.

Por este motivo, se mejoraron espectacularmente los sistemas de codificación de atención primaria (WONCA), hospitalaria (mínimum basic data set MBDS, o conjunto mínimo básico de datos, CMBD), y desde los años 1980's, los "diagnostic related groups" o grupos de diagnóstico relacionado (GRD), que permitieron por primera vez medir la actividad de manera analítica, más allá de la estancia media y ocupación media del centro. En Europa, hasta bien entrado el siglo XXI se estimaba que menos de la mitad de los hospitales tenían alguna métrica que les permitiera realizar la menor contabilidad analítica.

En Europa, hasta bien entrado el siglo XXI se estimaba que menos de la mitad de los hospitales tenían alguna métrica que les permitiera realizar la menor contabilidad analítica.

La cuestión moral aquí es una cuestión de eficiencia económica, y sin datos reales es cuestión de mera opinión. A todos estos desarrollos, Sandor Relman, por aquel entonces editor del New England Journal of Medicine le llamó, en 1988 "la tercera revolución sanitaria", a la de la transparencia y evaluación, habiendo sido la primera el progreso técnico espectacular, tanto en farmacología como en técnica, formación e infraestructuras de la postguerra mundial, y la segunda los mecanismos de control de costes, más o menos brutales de los años posteriores a la crisis del petróleo (16).

4.2 Síntesis horizontales y verticales. Otros intentos de medir los resultados en salud.

Las ciencias sociales ya habían puesto en circulación el concepto de utilidad, y se empezaba a difundir, con relativo entusiasmo, la esperanza de que los estudios coste-consecuencia podían aportar mucha racionalidad al sector sanitario.

A todos estos desarrollos, Sandor Relman, por aquel entonces editor del New England Journal of Medicine le llamó, en 1988 "la tercera revolución sanitaria", a la de la transparencia y evaluación

Se dio la coincidencia en el tiempo de distintas iniciativas que confluyeron.

A) Desde los trabajos de Donabedian de 1966 se empezó a distinguir en primer lugar los conceptos de estructura, proceso, y resultados. En dichos trabajos se consideraba que los resultados no eran el único parámetro, pero sí el más importante (17).

B) Había cierta experiencia en programas de desarrollo (Banco Mundial, OMS, programas de erradicación de transmisibles) acerca del concepto "burden of disease" (carga de enfermedad) y su cuantificación en términos económicos y sanitarios, lo que daba cierta credibilidad a que fuere posible, e incluso deseable, llegar al análisis coste-beneficio robusto, paradigma del buen gobierno y de la transparencia.

C) Se habían desarrollado los primeros conceptos de medición de la salud de tipo genérico, multidimensionales, y se estaban validando activamente en diversos tipos de subpoblaciones, como el Nottingham Health Profile, y el Sickness Impact Profile, cuyos procesos de validación parecían sensibles, es decir, con un buen potencial de validez de constructo. Asimismo se estaban validando instrumentos para la medición de los estados de salud mental para la población general que empezaron a permitir, es decir, hacer viable, tener buena epidemiología psico-psiquiátrica, como el Goldberg Health Questionary (GHQ), simple de aplicar, y que mostró la alta prevalencia de estados de depresión leve, subclínica pero tratable y el "Present State Examination" de Sartorius y el instituto de psiquiatría de Londres que permitió por primera vez la existencia de datos consistentes internacionales en psicopatologías mayores.

D) Se desarrolló un estudio de dimensiones desconocidas hasta la época (10.000 pacientes) para su validación sobre el estado de salud para pacientes ambulatorios, el Medical Outcomes Study (MOS), del cual se obtuvieron cuestionarios reducidos razonablemente robustos: el Short-Form 36 (SF36), luego reducido a SF-24 e incluso SF-12.

Parecía que había llegado la época de oro de la investigación en servicios de

salud, y que iba a ser posible el juicio de Relman, transparencia gracias a datos razonablemente robustos, y racionalización de las decisiones, ya que iba a ser posible el análisis, no solo coste-efectividad, sino también coste-utilidad, y a nivel poblacional, el coste-beneficio.

4.3 Entrando en materia. Los PORT y la dura realidad.

Eran tiempos donde el entusiasmo por la MBE era creciente, y evidentemente, el punto crítico era la calidad de la evidencia. Parte del problema era, y sigue siendo, su fragmentación en estudios de entidad no demasiado grande. Una salida obvia era facilitar la síntesis, como se ha comentado, y de aquí surgió la colaboración Cochrane de Oxford, cuyo germen fue la primera base de datos de ensayos clínicos obstétrico-ginecológicos. Pero como hizo fortuna la frase: “el tamaño importa”, Es fundamental poder hacer análisis por subgrupos. Tiempos de “think big”. Uno de los artículos más famosos de la época abogaba por vincular la investigación clínica a la política sanitaria, y concentrar los esfuerzos en ensayos clínicos pragmáticos y “grandes” (18).

Esta propuesta no cayó en saco roto, y la administración Bush lanzó en 1988 una incitativa a priori deslumbrante: la creación de la “Agency for Healthcare Policy and Resarch” (AHPR), uno de cuyos principales mandatos era la puesta en marcha de una serie de macoproyectos de investigación llamados “Patient Outcomes Research Teams” (PORT), temáticos. Cada PORT debía enfocar una patología prevalente y significativa, desarrollar y validar medidas de resultados sensibles e inteligibles para el paciente (“patient reported outcome”), realizar una síntesis de información del estado del arte, hacer un análisis detallado de las bases de datos administrativas de utilización y financiación (Medicare, aseguradoras) y realizar un ensayo clínico entre opciones terapéuticas reales. El slogan era “translating research into practice”, con toda una estrategia asociada.

El mandato de la AHPR fue renovado, y los resultados de estos proyectos evaluados (19). De alguna manera la realidad resultó ser más rebelde a lo que se esperaba, pero los progresos en saber lo que daba de sí la “efectiveness research” fueron notables, habiendo surgido la primera generación de investigadores en servicios sanitarios sólidos. Hubo una segunda convocatoria, la PORT-II y de alguna manera el abordaje había perdido frescura (20). Se empezaron a barajar conceptos aparentemente novedosos pero muy de gestión, tras el fracaso del intento de reforma sanitaria Clinton, como el de “disease management”, que en la práctica no han cuajado.

Una de las aportaciones de la época fue descubrir la tremenda iatrogenia oculta, resultando prioritario el abordaje de los temas de seguridad en el paciente.

El debate académico, tanto conceptual como instrumental sigue con vigor, con datos experimentales. Sabemos mucho más que hace una década acerca de las utilidades. Las herramientas existen, están validadas y los antiguos debates entre las métricas específicas de enfermedad y métricas genéricas hoy suenan a obsoletos.

5.- El estado actual de la cuestión

Evidentemente el progreso ha sido considerable, y la cuestión de la evidencia se ha vuelto indiscutible. El debate académico, tanto conceptual como instrumental sigue con vigor, con datos experimentales. Sabemos mucho más que hace una década acerca de las utilidades. Las herramientas existen, están validadas y los antiguos debates entre las métricas específicas de enfermedad y métricas genéricas hoy suenan a obsoletos. Sabemos que hay límites en la posibilidad de medir, y que el desarrollo de instrumentos tiene unos esfuerzos y unos costes ocultos que podemos estimar realísticamente.

En el escenario regulatorio que vivimos la cuestión del tiempo es clave por

tanto el problema de los “proxys” no tiene solución. La innovación implica riesgo, y la innovación sanitaria implica riesgos sanitarios. No se pueden plantear “endpoints” a unos plazos que sean médicamente sólidos, digamos más de 5 años, por lo tanto hay que usar muy a menudo variables substitutorias, que siempre son opciones no completamente satisfactorias.

No se pueden plantear “endpoints” a unos plazos que sean médicamente sólidos, digamos más de 5 años, por lo tanto hay que usar muy a menudo variables substitutorias, que siempre son opciones no completamente satisfactorias.

La primera y principal conclusión es que éste no es un tema para generalistas ni aficionados. Los estándares son muy altos, al igual que el rigor, para el farmacólogo clínico, para el epidemiólogo, para el monitor de los estudios, para el estadístico, y que dada la especificidad de la medición de outcomes hay muy pocos grupos capaces de desarrollar, validar y poner en valor un cuestionario “state of the art”. Lo mismo se aplica a los responsables de seguimiento, vigilancia postventa, márketing, y a cualquiera que participe en el equipo de forma no-pasiva.

Otra consideración adicional es la naturaleza competitiva de la oferta sanitaria. El legado de los PORTS es inconcebible desde el sector privado, que debe operar con su natural reserva, sea el farmacéutico o el asistencial donde éste está liberalizado. No debemos subestimar este legado: ha dado la base de solidez a la farmacoeconomía, que, no olvidemos, tiene dos puntales, uno en ciencias sociales y el otro en las experimentales, que son paradigmas epistemológicos distintos, con distintos estándares de robustez. Aquí nos encontramos con un punto problemático de difícil solución. Se compete en precio, o mejor dicho, en margen, o en ROI si vamos al “bottomline”. Pero el discernimiento entre necesidad, demanda y utilización implica entender que no están al mismo nivel. Hagamos el ejercicio de usar dos tipos de lenguaje, el de la necesidad y el de la demanda:

Necesidad	Demanda
Necesidades normativas	
Necesidades percibidas	Necesidad
Necesidades expresadas	Demanda real
Necesidades satisfechas	Utilización

Para el sector privado interesa que la información real de demanda sea propietaria, ya que parte de la competitividad es pasar de la demanda percibida pero no expresada se transforme en demanda real, y dentro del esquema de la gestión logística y gestión del margen (el esquema de Porter), para lo cual hay una miríada de iniciativas innovadoras de análisis refinado de datos, el llamado “data mining” que es propietario. La información sobre el RWD, el “real world data” es de muy difícil acceso al lego o profano, y elude la prensa académica, encontrándose poca información fragmentaria o en blogs, newsletters y similares.

Acabemos con un caso real. La culpa la tuvieron dos incidentes, uno de ellos reversible, que tuvieron lugar a los dos meses del lanzamiento de un “mirlo de oro”, es decir, un fármaco innovador, sin competencia a 2-3 años vista. Fármaco que justo había sido registrado en Londres para toda la UE y que solo había sido lanzado en dos mercados, perdón, dos países. Estos incidentes, concretamente de tipo ictericia por presunta hepatotoxicidad, de la que solo 1 paciente tuvo secuelas implicaron la caída en valor de la empresa en cuestión de días de un 35% de su valor de mercado bursátil a escala mundial, y tardó muchos meses en empezar a recuperarse. Era la primera antitrombina oral, que obviaba todos los inconvenientes de los anticoagulantes antivitamina K, la empresa, una líder en el sector, y había hecho 6 ensayos clínicos para diversas indicaciones, con un total acumulado de más de 20.000 pacientes.

Bibliografía

1. Moynihan R. What is disease? And why it's a healthy question. *BMJ*. 2013;346:f107.
2. Leavell H, Clark A. Preventive medicine for the doctors in his community. New York: McGraw-Hill; 1965.
3. Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease mongering. *BMJ*. 2002 Apr 13;324(7342):886-91.
4. Jeckell J. El Prof Jeckell fue el director del programa de evaluación sanitaria del Yale Department of Epidemiology and Public Health (1978-90).
5. Fletcher R, Fletcher S, Wagner E. *Clinical epidemiology: The Essentials*. Baltimore: Williams & Wilkins; 1982.
6. U.S. Preventive Services Task Force (USPSTF). U.S. Department of Health & Human Services; Accesible desde: <http://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/uspstf.html>
7. Garcia Lopez FJ, Gutierrez Bezon S, Galende Dominguez I, Avendano Sola C. [The evaluation of the quality of clinical trials: the rationale, usefulness and drawbacks]. *Medicina clinica*. 1999;112 Suppl 1:35-42.
8. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996 Feb;17(1):1-12.
9. CONSORT Statement Website. Accesible desde: <http://www.consort-statement.org>.
10. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007 Nov;18(6):805-35.
11. Knottnerus JA, Tugwell P. The standards for reporting of diagnostic accuracy. *J Clin Epidemiol*. 2003 Nov;56(11):1029.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009 Jul 21;6(7):e1000100.
13. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the strengthening the reporting of observational studies in epidemiology (STROBE) statement. *J Clin Epidemiol*. 2009 Jun;62(6):597-608 e4.
14. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication guidelines for improvement studies in health care: evolution of the SQUIRE Project. *Ann Intern Med*. 2008 Nov 4;149(9):670-6.
15. Gol-Freixa J. Bienvenidos a la medicina basada en la evidencia. *La medicina basada en la evidencia Guías del usuario de la literatura médica: JAMA*; 1997.
16. Relman AS. Assessment and accountability: the third revolution in medical care. *N Engl J Med*. 1988 Nov 3;319(18):1220-2.
17. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966 Jul;44(3):Suppl:166-206.
18. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984 Oct-Dec;3(4):409-22.
19. 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.
20. Stryer D, Tunis S, Hubbard H, Clancy C. The outcomes of outcomes and effectiveness research: impacts and lessons from the first decade. *Health services research*. 2000 Dec;35(5 Pt 1):977-93.



Dra. Montserrat Ferrer

Unidad de Investigación en Servicios Sanitarios. Instituto Municipal de Investigación Médica (IMIM). Barcelona

MEDIDAS DE SALUD REPORTADAS POR LOS PACIENTES: METODOLOGÍA, EMPLEO Y CORRELACIÓN CON INDICADORES CLÍNICOS.

1. DEFINICIÓN DE RESULTADOS REPORTADOS POR LOS PACIENTES.

Los resultados reportados o percibidos por los pacientes (del término en inglés Patient Reported Outcomes, PRO) es un concepto que cubre un rango amplio de diferentes tipos de resultados en salud (1) tales como síntomas, limitaciones en la actividad, estado de salud, calidad de vida relacionada con la salud, satisfacción del paciente o conformidad con el tratamiento. La característica común entre los distintos constructos agrupados es que recogen información directamente del paciente sin interpretación alguna por parte de médicos u otros profesionales sanitarios (2;3). No deben confundirse con escalas de puntuaciones clínicas, donde el especialista completa un formulario para puntuar la gravedad de la enfermedad o los efectos del tratamiento. En este sentido, la definición de la Food and Drug Administration (4) es muy explícita: "Un PRO es cualquier informe del estado de salud de un paciente, que proviene directamente del paciente, sin interpretación de la respuesta del paciente por parte de un médico o cualquier otra persona".

El término PRO fue acuñado aproximadamente en el año 2000, específicamente para evitar la confusión suscitada alrededor del término “calidad de vida”. La calidad de vida relacionada con la salud es un tipo o subgrupo específico dentro de los resultados percibidos por los pacientes, que se distingue por su multidimensionalidad. Aunque existen múltiples definiciones, todas tienen un sustrato común con la definición de salud proporcionada por la Organización Mundial de la Salud en 1948: “La salud es el estado de completo bienestar físico, psíquico y social de un individuo y no solo la ausencia de enfermedad”.

La calidad de vida relacionada con la salud se distingue por su multidimensionalidad

2. CARACTERÍSTICAS FUNDAMENTALES DE LOS INSTRUMENTOS

Este apartado resume brevemente las principales características que deben ser tenidas en cuenta al desarrollar, adaptar o evaluar un instrumento de medida de resultados percibidos por los pacientes. Para su elaboración hemos seguido la propuesta del Comité Científico Asesor del Medical Outcomes Trust (5), que ha sido la base para herramientas como el EMPRO (6) y el COSMIN (7) diseñadas, respectivamente, para valorar estos instrumentos y la calidad metodológica de los estudios realizados sobre ellos.

2.1. Concepto y modelo de medida

El modelo conceptual se puede definir como la justificación y descripción del concepto medido y de las poblaciones que el instrumento pretende evaluar. El modelo de medida es la forma en que se operacionaliza el modelo conceptual durante la fase de desarrollo del instrumento, que queda reflejado o expuesto en los métodos seguidos para obtener el pool inicial de ítems, seleccionarlos y combinarlos en dimensiones.

Entre los criterios de calidad que se deben considerar destacan: la participación de la población a la que va dirigido el instrumento en la obtención del contenido (grupos focales, entrevistas semi-estructuradas, prueba piloto, etc.), la evaluación de la dimensionalidad (mediante análisis factorial o teoría de respuesta al ítem, entre otros) y de las puntuaciones.

2.2. Fiabilidad

Se puede definir la fiabilidad como el grado en que un instrumento está libre de error aleatorio. Existen dos aproximaciones para evaluar la fiabilidad: la consistencia interna y la reproducibilidad. La consistencia interna es una estimación de la precisión del instrumento basada en la homogeneidad de los ítems en una única administración. El estimador de la consistencia interna en la psicometría clásica es el coeficiente alfa de Cronbach. La reproducibilidad o repetibilidad puede evaluarse a partir de la concordancia de los resultados obtenidos en diferentes administraciones, mediante diseños test-retest en individuos estables. El estimador de la concordancia en variables continuas es el coeficiente de correlación intraclase.

Para los coeficientes de fiabilidad (tanto alfa de Cronbach como coeficiente de correlación intraclase) se recomiendan valores de 0,7 o superiores para considerar un instrumento fiable a nivel de grupos y de 0,9 para el uso a nivel individual (comparar diferentes mediciones de un mismo individuo o diferentes individuos entre sí). En los estudios test-retest es importante considerar el intervalo entre las dos administraciones y cómo se ha determinado que el estado de salud no ha cambiado para justificar la asunción de estabilidad de la muestra.

Se puede definir la fiabilidad como el grado en que un instrumento está libre de error aleatorio

2.3. Validez

Se define como la capacidad del instrumento para medir lo que se pretende medir, y se distinguen tres tipos de validez según la forma en que se valora: de contenido, de constructo y de criterio. La validez de contenido valora la adecuación de los ítems y dimensiones del instrumento al propósito para el que ha sido diseñado mediante técnicas cualitativas (como paneles de expertos o grupos focales con pacientes) o cuantitativas (análisis factorial exploratorio y confirmatorio). La validez de constructo se evalúa mediante la demostración de los patrones de puntuación esperados en grupos conocidos (por ejemplo, según la clasificación de gravedad de una patología); y de las relaciones lógicas con otras medidas, como variables clínicas, de laboratorio, u otros instrumentos de resultados percibidos por los pacientes. En ambos casos es muy importante especificar las hipótesis a priori. La validez de criterio hace referencia a la asociación con un criterio de referencia o gold standard, que en este campo únicamente es posible cuando se evalúan versiones cortas de instrumentos preexistentes.

La capacidad del instrumento para medir lo que se pretende medir

2.4. Sensibilidad al cambio

La sensibilidad al cambio puede definirse como la capacidad del instrumento para detectar cambios en el tiempo. Para su evaluación debe estimarse la magnitud del cambio observado en estudios longitudinales en un grupo en el que hipotetizamos un cambio. Es importante estimar la magnitud del cambio con indicadores estandarizados (8) como el tamaño del efecto o el error estándar de la medida, comparar con el cambio experimentado por un grupo estable, justificar el intervalo entre administraciones, las intervenciones terapéuticas que producen el cambio

La sensibilidad al cambio puede definirse como la capacidad del instrumento para detectar cambios en el tiempo.

y cómo se ha determinado que el estado de salud ha cambiado o ha estado estable.

2.5. Interpretabilidad

Se define como interpretabilidad el grado en que es posible asignar un significado comprensible a las puntuaciones del instrumento. Los métodos más utilizados son: las normas de referencia, las estrategias basadas en "anclas" y los métodos basados en distribuciones como el tamaño del efecto.

A pesar de las críticas recibidas (9), la pregunta global de cambio ha sido el ancla más utilizada para estimar la 'Mínima Diferencia Importante' mediante la comparación de la diferencia en las puntuaciones entre dos administraciones con el cambio mínimo detectado por el paciente (10).

En el caso del tamaño del efecto, diferencias en las puntuaciones de 0,8, 0,5 y 0,2 desviaciones estándar han sido definidas (11), respectivamente, como cambios grandes, moderados y pequeños.

Se define como interpretabilidad el grado en que es posible asignar un significado comprensible a las puntuaciones del instrumento

2.6. Facilidad de uso del instrumento

La carga de administración se define como el tiempo, el esfuerzo, y el consumo de otros recursos que recaen sobre los individuos que responden o sobre los que lo administran. En este sentido es importante disponer de información sobre el tiempo necesario para completar el instrumento, el nivel de comprensión de lectura mínimo, y la complejidad del proceso de obtención de puntuaciones.

2.7. Formas de administración

Las características descritas (fiabilidad, validez, sensibilidad al cambio, interpretabilidad y facilidad de uso) deben ser evaluadas, no sólo en la versión original del instrumento, sino en cada una de las formas alternativas de administración que se propongan. Las formas alternativas deberían compararse con la original para realizar valoraciones de su equivalencia en las características métricas y en las puntuaciones obtenidas.

La carga de administración se define como el tiempo, el esfuerzo, y el consumo de otros recursos que recaen sobre los individuos que responden

2.8. Proceso de adaptación cultural y lingüística

La adaptación cultural del instrumento se refiere al desarrollo de una nueva versión equivalente a la original al ser administrada a una población diferente de la que fue diseñado. El proceso de adaptación lingüística debe incluir por lo menos dos traducciones independientes de la lengua de origen a la lengua que se quiere adaptar, una (preferiblemente más) traducción inversa a la lengua de origen, revisión de las versiones traducidas por paneles de profesionales y evaluaciones con pacientes (o población a la que vaya dirigido) mediante técnicas cualitativas para proporcionar evidencia de la aceptabilidad y comprensión.

La adaptación cultural del instrumento se refiere al desarrollo de una nueva versión equivalente a la original

Es importante la evaluación de la equivalencia conceptual para lograr la identificación de cualquier diferencia relevante entre el original y la versión adaptada, con argumentación para explicar estas diferencias e información sobre cómo fue llevada a cabo la reconciliación de las incongruencias. De forma similar

a la valoración de las formas alternativas de administración, las características del instrumento (fiabilidad, validez, sensibilidad al cambio, interpretabilidad y facilidad de uso) deben ser evaluadas en cada una de las nuevas versiones adaptadas.

3. APLICACIONES DE LAS MEDIDAS DE SALUD REPORTADAS POR LOS PACIENTES.

Durante la segunda mitad del siglo pasado se han ido acumulando las evidencias científicas sobre la calidad de los instrumentos de medida de los resultados percibidos por los pacientes y su importante poder predictivo de la mortalidad, la demanda de servicios y los costes sanitarios. Sin embargo, la incorporación de los resultados percibidos por los pacientes a la investigación, la práctica clínica o la planificación sanitaria no ha sido un proceso sencillo. Algunas de las barreras que han dificultado que estas medidas tengan un papel relevante en la toma de decisiones sanitarias, son la falta de familiaridad que dificulta su interpretación, o la carga adicional que supone su aplicación.

La tabla recoge la lista de las once aplicaciones que han sido documentadas para los resultados percibidos por los pacientes (5). Estas medidas se han ido incorporando a las encuestas nacionales de salud de diversos países (como Reino Unido (12), Finlandia (13), Suecia (14) o EEUU (15) entre otros), demostrando su utilidad para la evaluación y monitorización de la salud de la población general. También han resultado útiles en la evaluación del impacto de variadas intervenciones de salud, como por ejemplo la ley antitabaco de 2006 en Escocia (16). Seguramente una de sus aplicaciones más clásicas en investigación ha sido en los ensayos clínicos (17) para evaluar la eficacia y efectividad de las intervenciones. La reciente publicación de la extensión de la guía CONSORT (18), así como las recomendaciones de la FDA (Food and Drug Administration) (4) y de la EMEA (European Medicines Agency) (19) para incorporar información sobre calidad de vida en las fichas técnicas de los medicamentos, son pruebas evidentes del interés creciente suscitado en este ámbito.

En el Reino Unido, la medición de resultados percibidos por los pacientes ha experimentado recientemente un impulso definitivo tanto con las evaluaciones económicas del NICE (National Institute for Health and Clinical Excellence) con el Euroqol, como con la iniciativa PROMS (20) del National Health Service (NHS) para el análisis de la calidad asistencial. Desde 2009, se recogen mediciones de resultados percibidos en todos los pacientes sometidos con financiación del NHS a cuatro procedimientos quirúrgicos electivos (hernia inguinal, varices, prótesis de cadera y rodilla).

La detección precoz de problemas de salud mediante este tipo de instrumentos ha sido más aplicada en el campo de la salud mental, en el que han demostrado ser útiles en el proceso diagnóstico de la depresión (21;22) y otros trastornos psiquiátricos. Finalmente, pero no menos importante, este tipo de medidas se están incorporando a la práctica clínica habitual (23). La monitorización del estado de salud individual tiene un potencial importante en la mejora del manejo clínico del paciente y la provisión de los servicios sanitarios pertinentes. Para cada una de las aplicaciones comentadas necesitaremos utilizar un instrumento adecuado para tal fin. En los subapartados siguientes se describen brevemente los principales tipos de instrumentos y los recursos disponibles para identificar y seleccionar los instrumentos.

Para cada una de las aplicaciones comentadas necesitaremos utilizar un instrumento adecuado para tal fin.

3.1. Tipos de instrumentos de resultados percibidos por los pacientes

Independientemente del constructo medido, se pueden diferenciar dos grandes grupos de instrumentos según la población a la que se dirigen, los genéricos y los específicos. Los instrumentos genéricos están diseñados para ser utilizados en todo tipo de individuos, independientemente de la patología o síntomas que presente, y permiten las comparaciones entre grupos

de pacientes con distinta enfermedad. En contrapartida, los cuestionarios específicos están diseñados para pacientes con una determinada enfermedad, síntoma o población y son mejor aceptados porque incluyen sólo aquellas dimensiones relevantes para su patología, y pueden ser más sensibles en la detección de los efectos de las intervenciones terapéuticas. Una aproximación adecuada en proyectos de investigación es incluir un instrumento genérico y otro específico.

Los instrumentos genéricos están diseñados para ser utilizados en todo tipo de individuos; los cuestionarios específicos están diseñados para pacientes con una determinada enfermedad

3.2. Recursos para su identificación, selección y uso

El desarrollo de las medidas de Resultados Percibidos por los Pacientes ha seguido una tendencia prácticamente exponencial en las últimas décadas. La gran cantidad de instrumentos existentes en la actualidad hace difícil identificar los instrumentos disponibles para un uso o una población específica.

Actualmente, la solución más rápida y más eficaz son las bases de datos de instrumentos. Quizás la más importante es la base de datos PROQOLID (<http://www.proqolid.org/>), que dispone de información de instrumentos para todo tipo de poblaciones y patologías, aunque sólo es de libre acceso una parte de la información. También existen diversas bases de datos especializadas en patologías o especialidades médicas concretas, algunas de ellas alojadas en la página web de sus sociedades científicas.

El desarrollo de las medidas de Resultados Percibidos por los Pacientes ha seguido una tendencia prácticamente exponencial en las últimas décadas

No obstante, estas bases de datos no se han especializado en los instrumentos en español. Para obtener información sobre cuestionarios españoles, cabe

destacar la librería virtual BiblioPRO (<http://www.bibliopro.org/>), que pretende ser exhaustiva de los instrumentos de medida de Resultados Percibidos por los Pacientes en nuestro ámbito. Otra dirección a tener en cuenta cuando se buscan instrumentos en castellano es la web de BIPFAES, un Banco de Instrumentos Psicométricos (<http://bipfaes.faes.es/>). Una consulta realizada a estas librerías virtuales antes de iniciar un nuevo proyecto puede ser de gran utilidad.

4. CORRELACIÓN CON INDICADORES CLÍNICOS

En general, las relaciones entre las variables clínicas tradicionales y las medidas del estado de salud no han sido conceptualizadas apropiadamente. El modelo de Wilson y Cleary (24) diferencia los resultados en salud según los conceptos subyacentes y propone relaciones causales específicas entre ellos. Situa las diferentes medidas de salud en un continuum de complejidad biológica, social y psicológica creciente. En un extremo se encuentran las medidas biológicas, tales como los niveles de albúmina en suero y hematocrito, y en el otro las más integradas y complejas, tales como la calidad de vida y la salud general percibida.

De acuerdo con el modelo causal propuesto por estos autores (24), la asociación de la calidad de vida relacionada con la salud sería más baja con los indicadores biológicos o fisiológicos por estar situados en los extremos del continuo, intermedios con los síntomas y más altos con la capacidad funcional, por ser este el indicador más próximo. Existen evidencias en diferentes patologías que apoyan este gradiente. Por ejemplo, en los pacientes con insuficiencia cardiaca, las correlaciones de las puntuaciones de calidad de vida relacionada con la salud son bajas (inferiores a 0,3) con la fracción de eyección ventricular (25) y moderadas (alrededor de 0,5-0,6) con indicadores de capacidad funcional (26) como la distancia caminada en 6 minutos o la clasificación de la New York Heart Association. De forma similar, en los pacientes con enfermedad pulmonar obstructiva crónica (27), las

Las relaciones entre las variables clínicas tradicionales y las medidas del estado de salud no han sido conceptualizadas apropiadamente

correlaciones de las puntuaciones de calidad de vida relacionada con la salud son bajas (inferiores a 0,3) con el volumen espiratorio forzado en el primer segundo (FEV1) y moderadas (alrededor de 0,5) con la disnea.

Debido a esta baja o moderada asociación entre indicadores clínicos tradicionales y los resultados en salud percibidos por los pacientes, los estudios de evaluación de intervenciones pueden aportar evidencias divergentes (utilizando unos u otros) y complementarias. Las medidas de calidad de vida relacionada con la salud pueden mostrar el beneficio de una intervención terapéutica cuando no se encuentran cambios en las medidas biológicas o fisiológicas, permitiendo que los pacientes se beneficien de un tratamiento efectivo que de otra forma hubiera sido valorado como inefectivo. Por ejemplo, en la rehabilitación respiratoria se detecte beneficio en términos de calidad de vida relacionada con la salud y de tolerancia al ejercicio (28). También puede ocurrir que exista mejoría en las variables tradicionales aunque no se acompañe de mejoría percibida por los pacientes, como en el caso de la oxigenoterapia domiciliar que ha mostrado beneficio en supervivencia (29). En este caso, la información aportada por las medidas de resultado percibido por los pacientes puede ser importante en el momento de plantear el tratamiento al enfermo y aumentar el grado de cumplimiento.

TABLA1. APLICACIONES DOCUMENTADAS (5) PARA LAS MEDIDAS DE RESULTADOS PERCIBIDOS POR LOS PACIENTES.

1	Evaluación transversal de la salud de la población general
2	Evaluación transversal de la salud de poblaciones específicas
3	Monitorización longitudinal de la salud de la población general
4	Monitorización longitudinal de la salud de poblaciones específicas
5	Evaluación del impacto de intervenciones o políticas sanitarias
6	Evaluación de la eficacia y efectividad de las intervenciones
7	Evaluación económica de las intervenciones sanitarias
8	Análisis de la calidad de las intervenciones sanitarias
9	Detección precoz de problemas de salud
10	Diagnóstico de problemas de salud
11	Monitorización del estado de salud individual

Bibliografía

1. Mckenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med.* 2011;9:86.
2. Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health.* 2003 Sep;6(5):522-31.
3. Doward LC, Mckenna SP. Defining patient-reported outcomes. *Value Health.* 2004 Sep;7 Suppl 1:S4-S8.
4. FDA. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville: Food and Drug Administration (FDA). Available from: <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> [Accessed February 20, 2013]; 2009.
5. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002 May;11(3):193-205.
6. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures. *Value Health.* 2008 Jan 8;11(4):700-8.
7. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012 May;21(4):651-7.
8. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials.* 1991;12:142S-58S.
9. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care.* 2001 Oct;39(10):1039-47.
10. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract.* 2000 Feb;6(1):39-49.
11. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
12. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ.* 1998 Mar 7;316(7133):736-41.
13. Saarni SI, Harkanen T, Sintonen H, Suvisaari J, Koskinen S, Aromaa A, et al. The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. *Qual Life Res.* 2006 Oct;15(8):1403-14.
14. Burstrom K, Johannesson M, Rehnberg C. Deteriorating health status in Stockholm 1998-2002: results from repeated population surveys using the EQ-5D. *Qual Life Res.* 2007 Nov;16(9):1547-53.
15. Luo N, Johnson JA, Shaw JW, Feeny D, Coons SJ. Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Med Care.* 2005 Nov;43(11):1078-86.
16. Menzies D, Nair A, Williamson PA, Schembri S, Al-Khairalla MZ, Barnes M, et al. Respiratory symptoms, pulmonary function, and markers of inflammation among bar workers before and after a legislative ban on smoking in public places. *JAMA.* 2006 Oct 11;296(14):1742-8.
17. Soto J, Rejas J, Ruiz M, Pardo A. [Health-related quality of life studies in clinical trials: Assessment of their methodological quality through a check-list.]. *Med Clin (Barc).* 2009 Sep 5;133(8):314-20.
18. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013 Feb 27;309(8):814-.
19. EMEA. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. London: EMEA; 2006.
20. Black N. Patient reported outcome measures could help transform healthcare. *BMJ.* 2013;346:f167.

[48] Medidas de salud reportadas por los pacientes: metodología, empleo y correlación con indicadores clínicos

21. Williams JW, Jr., Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA*. 2002 Mar 6;287(9):1160-70.
22. Gill SC, Butterworth P, Rodgers B, Mackinnon A. Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population. *Psychiatry Res*. 2007 Jul 30;152(1):63-71.
23. Santana MJ, Feeny DH. Using the Health Utilities Index in Routine Clinical Care: Process, Feasibility, and Acceptability. A Randomized Controlled Trial. *Patient*. 2009;2(3):159-67.
24. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273(1):59-65.
25. Juenger J, Schellberg D, Kraemer S, Haunstetter A, Zugck C, Herzog W, et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. *Heart*. 2002 Mar;87(3):235-41.
26. Garin O, Ferrer M, Pont A, Rue M, Kotzeva A, Wiklund I, et al. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. *Qual Life Res*. 2009 Feb;18:71-85.
27. Curtis JR, Deyo RA, Hudson LD. Health-related quality of life among patients with chronic obstructive pulmonary disease. *Thorax*. 1994;49:162-70.
28. Lacasse Y, Wong E, Guyatt GH, King D, Cook DJ, Goldstein RS. Meta-analysis of respiratory rehabilitation in chronic obstructive pulmonary disease. *Lancet*. 1996;348:1115-9.
29. Okubadejo AA, Paul EA, Jones PW, Wedzicha JA. Does long-term oxygen therapy affect quality of life in patients with chronic obstructive pulmonary disease and severe hypoxaemia? *Eur Respir J*. 1996;9(11):2335-9.



Dr. José María Abellán Perpiñán

Grupo de Trabajo de Economía de la Salud
(GTES), Universidad de Murcia

EL AVAC COMO INDICADOR DE LA ESPERANZA Y LA CALIDAD DE VIDA

Los Años de Vida Ajustados por la Calidad (AVAC) poseen una dilatada historia, pues su origen se remonta a los años setenta del pasado siglo (1). Desde entonces, lo simple e intuitivo de su cálculo contribuyó a que su utilización se popularizara rápidamente, constituyéndose con el transcurso del tiempo en la medida de resultados recomendada para el análisis coste-efectividad de las intervenciones sanitarias, tanto por paneles de expertos (2) como por instituciones. Así, el National Institute for Health and Clinical Excellence (NICE), la agencia de evaluación económica de tecnologías sanitarias más influyente del mundo, recomienda en sus directrices metodológicas (3), que: “Para análisis coste-efectividad, el valor de los efectos de salud deberían expresarse en términos de AVAC para el horizonte temporal apropiado.”

El atractivo de los AVAC frente a indicadores tradicionales de salud, como la esperanza de vida al nacer o la tasa de mortalidad, radica en que integran en un único número no solo la cantidad de vida del paciente, sino también su calidad de vida. Se erige así como una medida del output final del sector sanitario, ya que es capaz de reflejar de un modo integral las variaciones de morbi-mortalidad (variaciones en la esperanza de vida y en la calidad de vida) acaecidas a resultas de la aplicación de programas y tecnologías sanitarias. Además, en la medida que todos los tratamientos médicos originan cambios, bien en el tiempo de vida, bien en la calidad de vida relacionada con la salud (CVRS), bien en ambas dimensiones conjuntamente, los AVAC constituyen una unidad de medida homogénea que permite comparar entre sí todo tipo de intervenciones, lo cual justifica su gran utilidad para el análisis coste-efectividad. La costumbre hoy extendida de construir tablas clasificatorias (más conocidas por la expresión tomada directamente del inglés, league tables), comparando la razón incremental coste por AVAC de una nueva tecnología con las razones incrementales de otras tecnologías, encuentra su origen en esta facultad de los AVAC para facilitar la comparación de los resultados alcanzados por programas independientes (4).

Los AVAC integran en un único número no solo la cantidad de vida del paciente, sino también su calidad de vida

El AVAC puede interpretarse como un año en buena salud. En consecuencia, los AVAC derivados de una determinada intervención pueden concebirse como el número de años en buena salud equivalentes a la ganancia bruta de supervivencia procurada por aquella. Así, por ejemplo, los 5 años adicionales de vida que puede ganar un enfermo de cáncer a consecuencia de someterse a una laringectomía, equivaldrían a un número inferior de AVAC, toda vez que la CVRS del paciente sufre un deterioro (privación total o parcial del habla). El quid de

Los AVAC permiten comparar entre sí todo tipo de intervenciones

la cuestión radica en averiguar en cuánto se cifra dicha pérdida de calidad de vida o, dicho de otro modo, a qué fracción de un año de vida en buena salud (esto es, ajustado por la calidad) equivale un año de vida bruto (sin ajustar) ganado. La respuesta a esta pregunta es que depende, puesto que dicha fracción no se determina objetivamente, sino que es una valoración subjetiva que comúnmente se denomina 'peso de calidad de vida' o 'utilidad' (trasunto de bienestar), declarada por aquellas personas cuyas preferencias (sinónimo de satisfacción) se consideran como más relevantes al efecto de evaluar económicamente la intervención en cuestión. En definitiva, y aunque existen visiones discrepantes respecto de la interpretación que asumimos en este capítulo, el "ajuste por la calidad" implícito en el cálculo del los AVAC representa una valoración cardinal de las preferencias individuales¹, de modo que los mismos AVAC no dejan de ser una medida del bienestar o utilidad individual (5).

El "ajuste por la calidad" implícito en el cálculo del los AVAC representa una valoración cardinal de las preferencias individuales

Cálculo de los AVAC

La metodología y supuestos implícitos en el cómputo de los AVAC, difieren dependiendo de que la CVRS del paciente permanezca constante hasta su fallecimiento (perfil o estado de salud crónico) o bien, como es más realista suponer, varíe a lo largo del tiempo (perfil de salud variable).

En el primer caso, los AVAC son el producto de dos factores: los años de vida y la utilidad o bienestar asociado a la CVRS en la que se va a vivir durante esos años. Esta combinación la podemos representar de forma muy simple del siguiente modo:

(1) Hay algunos autores, como por ejemplo Culyer (1989), que conciben el AVAC como una medida de la salud de la población, no como una representación mensurable de las preferencias individuales. A esta perspectiva ajena a la interpretación convencional que se hace de los AVAC se la conoce como extra-bienestarismo (extra-welfarism).

Ecuación 1 $AVAC = U(Q, T) = V(Q) \times T$

T= años de vida; Q= calidad de vida relacionada con la salud; U(Q, T)= utilidad o bienestar derivado de vivir T años con la calidad de vida Q; V(Q)= utilidad o peso de calidad de vida asociada a Q.

La anterior ecuación puede extenderse a situaciones en las que la calidad de vida o estado de salud no permanece constante, sumando simplemente tantos productos $V(Q) \times T$ como variaciones experimente Q a lo largo del tiempo. Esto es, si fraccionamos el horizonte vital del paciente en tantos intervalos temporales disjuntos (no solapados) como ocasiones cambia la CVRS hasta producirse la muerte de aquel, el cálculo de los AVAC se reduce a:

La calidad de vida o estado de salud no permanece constante

Ecuación 2 $AVAC = \sum_{t=1}^n V(Q_t) \times T_t$

n= número de intervalos temporales disjuntos; t= periodo temporal.

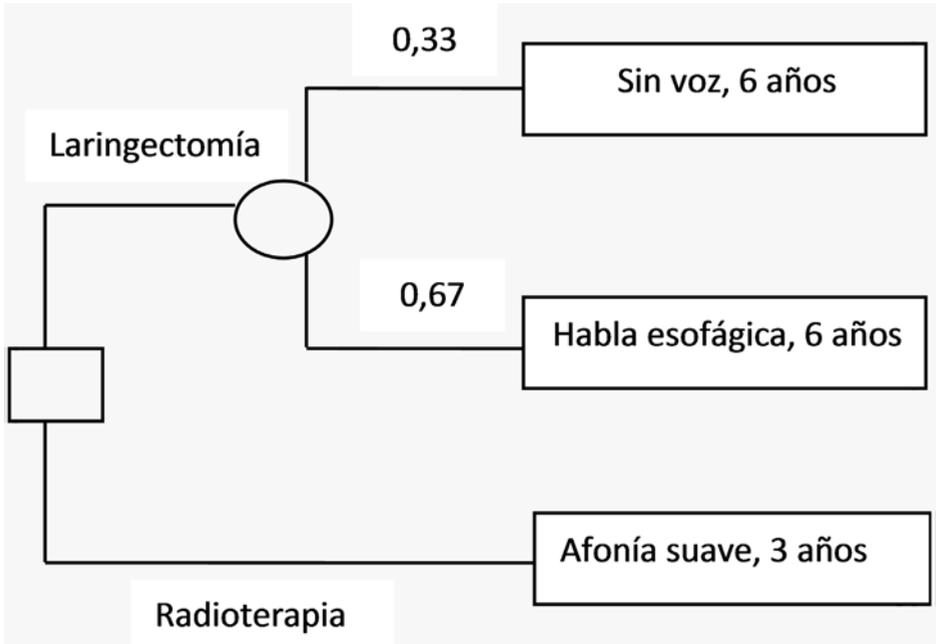
Podemos concretar en cuatro los pasos a seguir para poder comparar los AVAC producidos por dos intervenciones sanitarias mutuamente excluyentes:

1. Obtener la utilidad de la CVRS; esto es, medir $V(Q)$.
2. Multiplicar dicha utilidad por los años de vida T.
3. Calcular el número total de AVAC resultante, bien mediante la ecuación 1, bien mediante la ecuación 2, según que el perfil de salud sea constante o variable.
4. Hallar la diferencia entre los AVAC aportados por el nuevo programa y los AVAC aportados por el programa utilizado como comparador (la práctica habitual o status quo).

Un sencillo ejemplo ayudará a visualizar la aplicación de estos pasos.

Imaginemos que como alternativa a la laringectomía a la que antes aludíamos, existe la posibilidad de que el paciente se someta a radioterapia. Esta situación se representa por medio del árbol de Figura 1. Dos tratamientos alternativos para el cáncer de laringe

FIGURA 1. DOS TRATAMIENTOS ALTERNATIVOS PARA EL CÁNCER DE LARINGE



Fuente: elaboración propia a partir de Sox (1988) (6).

De acuerdo a los pasos enumerados, el primero de ellos (el más importante en realidad) consiste en la valoración de las preferencias del individuo; esto es, en la obtención de la utilidad de la CVRS. En el ejemplo concreto que nos ocupa, tres son los estados de salud a valorar (de mayor a menor gravedad): sin voz, habla esofágica y afonía suave.

Dos son las maneras de obtener las utilidades de la CVRS: la primera consiste en medir directamente las preferencias de la población objeto de interés (p.ej. un grupo de pacientes) con técnicas diseñadas al efecto; la segunda pasa por atribuir al estado de salud de la población diana las utilidades predichas por un

algoritmo o modelo de utilidad multiatributo estimado a partir de mediciones directas realizadas en amplias muestras de la población general.

Dentro de las medidas directas de preferencias, las que tradicionalmente han disfrutado de un mayor crédito entre los analistas son los instrumentos basados en elecciones, como es el caso de la lotería estándar (7) o el intercambio temporal (8), en contraposición a procedimientos que podríamos calificar como introspectivos, como es el caso de la escala visual analógica (9)². Otra forma de diferenciar los métodos directos de obtención de preferencias es atendiendo al tipo de decisión que evocan durante la medición; así, la lotería estándar estructura la decisión en un entorno de riesgo o incertidumbre, mientras que el intercambio de tiempos lo hace en un entorno de certeza. Asimismo, según la perspectiva suscitada en la medición, tenemos por un lado los métodos que participan de una perspectiva personal (todos los mencionados hasta el momento), mientras que otros procedimientos como el intercambio de personas(10) o el método del 'velo de la ignorancia' (11) presuntamente suscitan una perspectiva social que potencialmente al menos puede reflejar una cierta consideración por la equidad. Por ser las dos técnicas más extendidas, nos detendremos a continuación en explicar sucintamente cómo aplicar el método de la lotería estándar y los métodos del intercambio temporal para el caso de estados de salud crónicos, considerados mejores que la muerte (7)³.

La lotería estándar compara dos alternativas (figura 2). La alternativa 1 es un tratamiento con dos posibles resultados: o el paciente recobra su salud normal y vive durante t años más (probabilidad p) o bien muere (probabilidad $1-p$). El tratamiento incierto así descrito, por su semejanza con un juego de azar, recibe en esta literatura científica el nombre de 'lotería'. En contraposición, la alternativa

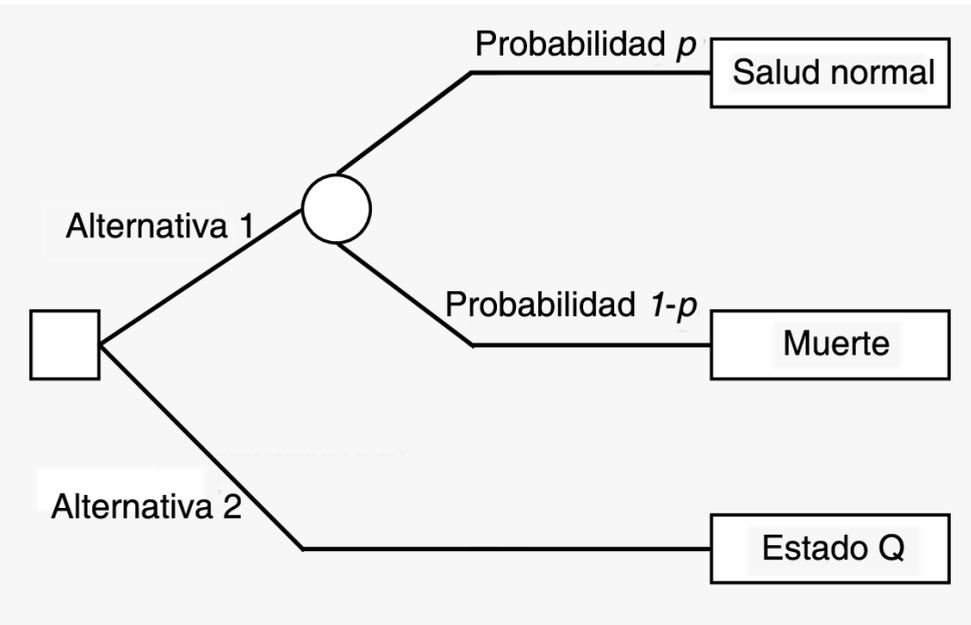
(2) Los métodos basados en elecciones se denominan así porque la preferencia es el resultado de la comparación entre dos alternativas. Por el contrario, la preferencia dimanante de técnicas como la escala visual analógica se fundamenta, en principio, en la introspección. Estos últimos métodos se consideran dotados de una menor validez (esto es, de una menor capacidad para medir realmente preferencias cardinales) ya que son muy sensibles a efectos contextuales (Torrance et al., 2001), si bien el debate acerca de la naturaleza y propiedades de la escala visual analógica continúa abierto.

(3) La preferencia por la permanencia en un determinado estado de salud no tiene por qué ser necesariamente monótona creciente con la duración de dicho estado (no hay nada que obligue a la persona de quien se obtienen las preferencias a preferir más a menos años en ese estado). El estado en cuestión puede percibirse como particularmente intolerable, de modo que el sujeto preferiría morir antes que vivir con dicha condición. Tanto la lotería estándar como el intercambio temporal pueden adaptarse para medir ese tipo de preferencias. Remitimos al lector interesado a Torrance (1986).

2 ofrece con certeza la seguridad de vivir en el estado de salud crónico Q durante t años. Partiendo de este planteamiento, el analista varía la probabilidad p hasta que el entrevistado se declare indiferente entre ambas alternativas, fijándose la utilidad de Q igual a p .

La lotería estándar compara dos alternativas (figura 2). La alternativa 1 es un tratamiento con dos posibles resultados: o el paciente recobra su salud normal y vive durante t años más (probabilidad p) o bien muere (probabilidad $1-p$). El tratamiento incierto así descrito, por su semejanza con un juego de azar, recibe en esta literatura científica el nombre de 'lotería'. En contraposición, la alternativa 2 ofrece con certeza la seguridad de vivir en el estado de salud crónico Q durante t años. Partiendo de este planteamiento, el analista varía la probabilidad p hasta que el entrevistado se declare indiferente entre ambas alternativas, fijándose la utilidad de Q igual a p .

FIGURA 2. EL MÉTODO DE LA LOTERÍA ESTÁNDAR

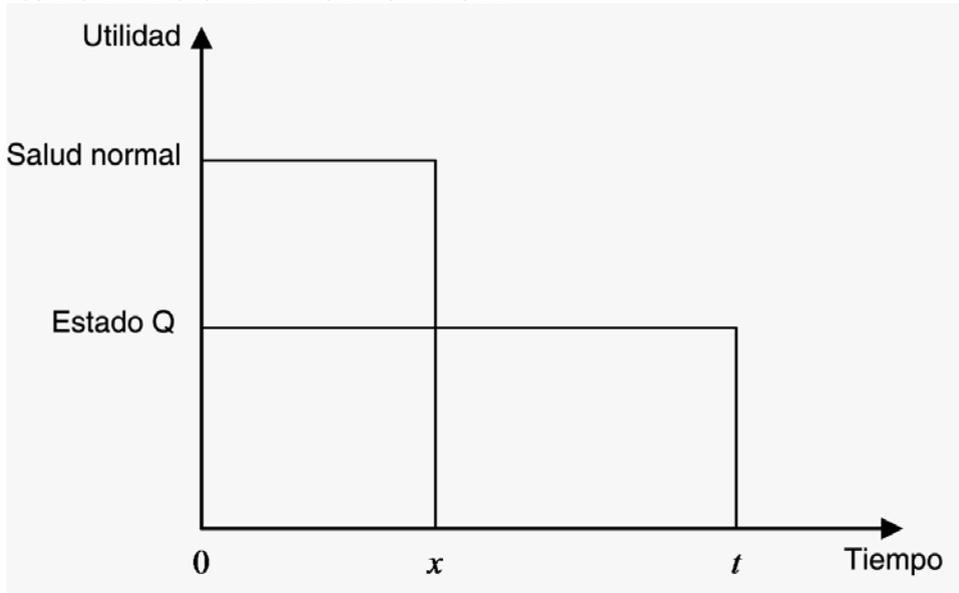


Fuente: elaboración propia a partir de Torrance (1986).

El intercambio temporal también plantea una elección entre dos alternativas

(figura 3). La alternativa 1 representa permanecer en el estado de salud Q (condición crónica considerada preferible a morir) durante t años, tras los cuales fallece, mientras que la alternativa 2 ofrece vivir x años ($x < t$) con salud normal. La duración x se va variando hasta que el individuo se muestra indiferente entre las dos alternativas, fijándose la utilidad de Q igual al cociente x/t (12)⁴.

FIGURA 3. EL MÉTODO DEL INTERCAMBIO TEMPORAL



Fuente: elaboración propia a partir de Torrance (1986).

Como alternativa a la medición directa de preferencias, pueden usarse las utilidades previamente estimadas por medio de instrumentos de utilidad multiatributo. Tres de los instrumentos multiatributo más ampliamente utilizados son el EuroQol-5D (13), el SF-6D (14) y el Health Utility Index (15). Estos instrumentos poseen dos componentes: un sistema de clasificación de estados de salud multidimensional y un algoritmo que asocia una utilidad a cada uno de los estados generados. Por ejemplo, el sistema EuroQol-5D consiste en la combinación de cinco dimensiones y tres niveles de gravedad por dimensión,

(4) Recientemente se ha propuesto un tipo de intercambio de tiempos diferente al tradicional explicado en esta página, denominado 'lead time tradeoff' (Devlin et al., 2011). La supuesta ventaja que ofrece esta variante de intercambio temporal es que permite obtener preferencias por estados de salud mejores y peores que la muerte de un modo uniforme, con el mismo formato.

de modo que puede describir hasta 243 (= 35) estados diferentes (16)⁵. A partir de estimaciones directas efectuadas con una muestra representativa de la población general, se ha estimado en varios países (en España por Badía et al., 2001(17)) el algoritmo del EQ-5D que asigna un valor a cada posible estado. Por su parte, el sistema clasificatorio del SF-6D es capaz de describir hasta 18.000 estados de salud diferentes (combinaciones de 6 dimensiones con un número variable de niveles, de 4 a 6, cada una de ellas). El algoritmo para España ha sido recientemente estimado por Abellán et al. (2012) (18). Por último, el HUI Mark 3 puede caracterizar casi 1.000.000 de estados de salud diferentes. Ruiz et al. (2003) estimaron dicho algoritmo para España (19).

Como alternativa a la medición directa de preferencias, pueden usarse las utilidades previamente estimadas por medio de instrumentos de utilidad multiatributo.

Ahora que ya sabemos cómo pueden obtenerse las utilidades de la CVRS, volvamos a nuestro ejemplo, y calculemos el número de AVAC asociados a cada uno de los dos tratamientos considerados, laringectomía y radioterapia. Supongamos que, por ejemplo, hemos obtenido la utilidad de cada uno de los tres estados de salud involucrados en la decisión representada en la figura 1 mediante el método del intercambio temporal. Imaginemos que el paciente a quien hemos valorados sus preferencias, ha declarado que considera equivalente pasar 5 años en buena salud a 10 en el estado 'sin habla', de modo que $V(\text{sin habla}) = 0,5$. Supongamos que de forma semejante, dicho paciente considera equivalente pasar 7,5 años en buena salud a vivir 10 años en el estado 'habla esofágica', y pasar 9,5 años en buena salud a vivir 10 en el estado 'afonía leve'. Las utilidades resultantes son, por tanto, 0,75 y 0,95, respectivamente. Ya tenemos las utilidades que multiplicadas por los años de vida correspondientes, constituyen los resultados de cada una de las ramas del árbol. Tal y como se recoge en la tabla 1, los AVAC así calculados son 3 (= $U(\text{sin habla}, 6 \text{ años})$), 4,5 (= $U(\text{habla esofágica}, 6 \text{ años})$) y 2,85 (= $U(\text{afonía leve}, 3 \text{ años})$). Si bien en el caso de la radioterapia, puesto que no hay riesgo, el número final de AVAC

(5) Recientemente se ha desarrollado una nueva versión del componente descriptivo del instrumento EQ-5D que consta de cinco niveles en lugar de tres (Herdman et al., 2011).

asociado al tratamiento son esos 2,85 AVAC que hemos mencionado, en el caso de la laringectomía, al ser un tratamiento arriesgado, hemos de combinar los resultados del mismo (los AVAC que obtendría el paciente a posteriori) con sus probabilidades de ocurrencia. Dicho de otro modo, para decidir ex ante qué tratamiento es más efectivo (produce más AVAC) hemos de estimar el número de AVAC que cabe esperar que gane el paciente si se somete a laringectomía. Para obtener dicho número esperado de AVAC, aplicamos la regla de evaluación de la teoría de la utilidad esperada (20), esto es:

$$\text{Ecuación 3: } AVAC \text{ esperados} = \sum_{i=1}^n p_i \times U_i(Q_i, T_i)$$

n = número de sucesos; *p_i* = probabilidad del suceso *i*; *U_i(Q_i, T_i)* = utilidad asociada al suceso *i*.

Aplicando esta ecuación a las probabilidades y utilidades asociadas a la laringectomía, obtenemos que dicho tratamiento produce un total de 4 AVAC (= 0,33x3 + 0,67x4,5 = 4,0), lo cual representa un incremento de 1,15 AVAC con respecto a la opción radioterapia (= 4,0 – 2,85 = 1,15).

TABLA 1. AVAC DE LOS TRATAMIENTOS LARINGECTOMÍA Y RADIOTERAPIA

	SIN HABLA	HABLA ES-OFÁGICA	AFONÍA LEVE	DIFERENCIA AVAC
V(Q)	0,5	0,75	0,95	
T	6	6	3	
U(Q, T)	3		2,85	
p	0,33	0,67	1	
p.U(Q, T)	0,99		2,85	
AVAC esperados	4,0		2,85	1,15

Condiciones de validez de los AVAC

Para que los AVAC representen fielmente las preferencias de la población, se requiere que dichas preferencias se ajusten a dos tipos de condiciones o supuestos:

- a) Supuestos generales que deben satisfacer las preferencias a consecuencia de la adopción de una determinada teoría de la utilidad.
- b) Supuestos específicos que caracterizan la fisonomía del modelo AVAC tal y como ha sido descrito por las ecuaciones 1 y 2.

Respecto a los supuestos generales que debe obedecer la preferencia por la salud para ser encarnada por el modelo AVAC, la mayoría de los trabajos teóricos existentes han caracterizado los AVAC como índices de utilidad esperada (21;22). La regla general ha sido, por tanto, suponer que las preferencias obedecen los axiomas de la teoría de la utilidad esperada.

En cuanto a los supuestos específicos que han de satisfacer las preferencias individuales para poder ser descritas mediante los AVAC, dos son las propiedades fundamentales que reúne la ecuación 1:

- a) Los dos atributos Q y T se combinan como un producto, propiedad ésta que es considerada como el rasgo intrínseco a los AVAC, e implica que la preferencia por ganar años de vida debe ser independiente de la preferencia por ganar calidad de vida, y viceversa.
- b) La utilidad del tiempo de vida T es lineal, razón por la cual la fórmula de cálculo de los AVAC se ha calificado como el modelo AVAC lineal. Esta condición, dentro del marco de la teoría de la utilidad esperada, requiere que el sujeto sea neutral al riesgo, esto es que sea indiferente entre un tratamiento incierto y un tratamiento seguro, produciendo ambos el mismo número esperado de años.

La extensión del modelo AVAC lineal al caso en que la CVRS varía a lo largo del

tiempo (ecuación 2) requiere que las preferencias satisfagan adicionalmente el supuesto de separabilidad o independencia aditiva (23). Este supuesto postula que una función de utilidad podrá descomponerse aditivamente si el individuo es indiferente entre dos loterías definidas sobre perfiles de salud variables del tipo $(Q_1, T_1; Q_2, T_2; \dots; Q_n, T_n)$ que tienen idénticas distribuciones de probabilidad sobre los diferentes niveles de los atributos.

Críticas y limitaciones de los AVAC

Probablemente la objeción más importante que se realiza a los AVAC radique en las dudas existentes respecto a la validez de sus supuestos. Por ejemplo, como ya sabemos, el hecho de que los AVAC se calculen como el producto de T y $V(Q)$ implica que se está suponiendo que la preferencia por el tiempo de vida es independiente de la preferencia por la calidad de vida, y viceversa. Este supuesto es contradictorio con preferencias como las que pueden producirse con pacientes con problemas de salud muy limitativos (p.ej. encamados), en las que a partir de un cierto periodo de tiempo ya no toleran vivir así (24)⁶. En casos como este los AVAC no describirán de forma válida las preferencias individuales. Otro ejemplo lo constituyen las evidencias recopiladas en contra del supuesto de linealidad de la utilidad del tiempo de vida (25).

Asimismo, si la teoría de la utilidad esperada falla al describir las preferencias individuales, y fundamentamos los AVAC sobre dicha teoría, entonces los AVAC tampoco serán válidos. Por desgracia hay abundante evidencia empírica que pone en cuestión la capacidad explicativa de la teoría de la utilidad esperada, lo cual ha conducido a intentar sustentar los AVAC en otro tipo de teorías alternativas (26). De hecho,

Probablemente la objeción más importante que se realiza a los AVAC radique en las dudas existentes respecto a la validez de sus supuestos

(6) Este fenómeno, observado empíricamente en varios estudios, se conoce como tiempo máximo tolerable (maximum endurable time) y fue descrito por vez primera por Sutherland et al. (1982).

durante años se ha considerado a la lotería estándar como la técnica de referencia (el gold standard), debido a que se deriva directamente de los axiomas de la teoría de la utilidad esperada. Sin embargo, y como han puesto de manifiesto diversos estudios, tanto la lotería estándar como el intercambio temporal están sujetos a sesgos (27) que ponen en cuestión la validez de las utilidades así obtenidas. Ante esta evidencia, hay dos posibles soluciones. Bien utilizar métodos de medición a priori menos sesgados, como, por ejemplo, los métodos de equivalencia entre loterías (18), bien aplicar correcciones cuantitativas a posteriori (debiasing) con el objeto de eliminar los sesgos (28).

Hay dos posibles soluciones: métodos de equivalencia entre loterías y correcciones cuantitativas a posteriori (debiasing)

Otro tipo de críticas provienen de la ausencia de las consideraciones de justicia distributiva al agregar los AVAC de la población. A este respecto, suele afirmarse que "1 AVAC vale 1 AVAC, no importa a quien corresponda". Hay que precisar ante esta objeción, no obstante, que nada impide introducir la dimensión equidad en el análisis coste-utilidad. Dicho de otra forma, los AVAC pueden ponderarse atendiendo a consideraciones de equidad (p.ej. la gravedad de la enfermedad), obteniendo así un determinado valor "social" del AVAC (29).

Bibliografía

1. Fanshel S, Bush J. A Health-Status Index and Its Application to Health-Services Outcomes. *Operations Research*. 1970;18:1021-66.
2. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.
3. NICE. Guide to the methods of technology appraisal 2008. Accesible desde: <http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf>.
4. Willinam A. Economics of coronary artery bypass grafting. *British medical journal (Clinical research ed)*. 1985 Sep 14;291(6497):736-7.
5. Culyer AJ. The normative economics of health care finance and provision. *Oxford Review of Economic Policy*. 1989;5:34-58.
6. Sox HC. *Medical Decision Making*. Butterworth-Heinemann; 1988.
7. Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of health economics*. 1986 Mar;5(1):1-30.
8. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health services research*. 1972 Summer;7(2):118-33.

9. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical decision making : an international journal of the Society for Medical Decision Making*. 2001 Jul-Aug;21(4):329-34.
10. Nord E. The person-trade-off approach to valuing health care programs. *Medical decision making : an international journal of the Society for Medical Decision Making*. 1995 Jul-Sep;15(3):201-8.
11. Pinto-Prades JL, Abellan-Perpignan JM. Measuring the health of populations: the veil of ignorance approach. *Health Econ*. 2005 Jan;14(1):69-82.
12. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Econ*. 2011 Mar;20(3):348-61.
13. Brooks R. EuroQol: the current state of play. *Health policy (Amsterdam, Netherlands)*. 1996 Jul;37(1):53-72.
14. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of health economics*. 2002 Mar;21(2):271-92.
15. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care*. 2002 Feb;40(2):113-28.
16. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2011 Dec;20(10):1727-36.
17. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2001 Jan-Feb;21(1):7-16.
18. Abellan Perpignan JM, Sanchez Martinez FI, Martinez Perez JE, Mendez I. Lowering the 'floor' of the SF-6D scoring algorithm using a lottery equivalent method. *Health Econ*. 2012 Nov;21(11):1271-85.
19. Ruiz M, Rejas J, Soto J, Pardo A, Rebollo I. [Adaptation and validation of the Health Utilities Index Mark 3 into Spanish and correction norms for Spanish population]. *Medicina clinica*. 2003 Feb 1;120(3):89-96.
20. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton NJ: Princeton University Press; 1944.
21. Pliskin J, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Research*. 1980;28(1):206-54.
22. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by Risk Neutrality. *Journal of Risk and Uncertainty* 1997;15:107-14.
23. Abellán J, Pinto J. Quality Adjusted Life Years as expected utilities. *Spanish Economic Review*. 2000;2(1):49-63.
24. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes toward quality of survival. The concept of "maximal endurable time". *Medical decision making : an international journal of the Society for Medical Decision Making*. 1982;2(3):299-309.
25. Abellán J, Pinto J, Ménendez-Martínez I, Badía X. Towards a better QALY model. *Health Economics*. 2006;15:665-76.
26. Abellan-Perpignan JM, Bleichrodt H, Pinto-Prades JL. The predictive validity of prospect theory versus expected utility in health utility measurement. *Journal of health economics*. 2009 Dec;28(6):1039-47.
27. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ*. 2002 Jul;11(5):447-56.
28. Pinto-Prades JL, Abellan Perpignan JM. When normative and descriptive diverge: how to bridge the difference. *Social Choice and Welfare* 2012;38:569-84.
29. Abellan Perpignan JM, Navarro J. Utilización de AVAC en diferentes situaciones clínicas. Sevilla: Informes de evaluación de tecnologías sanitarias AETSA 2009/10. Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía; 2011. Accesible desde: http://www.juntadeandalucia.es/salud/servicios/contenidos/nuevaaetsa/up/AETSA_2009-10_AVACs.pdf.



D. Isaac Aranda

Investigador

Seminario de Investigación en Economía y Salud UCLM

Profesor Ayudante en el área Fundamentos de Análisis Económico de la Universidad de Castilla la Mancha. Licenciado en Administración y Dirección de Empresas.

ASPECTOS DE INTERÉS PARA EL SNS

La medición de la calidad de vida relacionada con la salud comienza a ser una información que los decisores sanitarios persiguen obtener. La incorporación de escalas validadas en Encuestas oficiales sobre la salud de los ciudadanos muestran el elevado interés en identificar el efecto de las intervenciones sanitarias sobre el estado de salud de la población.

Entre los tipos de análisis que se pueden llevar a cabo en una evaluación económica, la calidad de vida relacionada con la salud (CVRS)¹ es utilizada como variable fundamental en los estudios de coste-utilidad. De hecho en nuestro país han aumentado el número de estudios que incluyan este tipo de análisis notablemente hasta alcanzar el 16% de las publicaciones realizadas en España durante el periodo 1983-2008 (1).

La relación entre la calidad de vida relacionada con la salud y el consumo de recursos está ampliamente estudiada (2-8). A pesar de los límites metodológicos de las escalas que persiguen identificar esta característica, es no obstante, un elemento recomendado por la guía española de elaboración de evaluaciones de tecnologías sanitarias cuando la intervención evaluada suponga a priori un cambio en el paciente que la utiliza. Esta guía recomienda que se informe de forma separada los años de vida ganados con las intervenciones comparadas o evaluadas y las tarifas que ajustan dicha ganancia según la calidad percibida por los pacientes. También apunta que si como consecuencia de la intervención o la enfermedad existen cuidados informales, la calidad de vida de los cuidadores debe aparecer diferenciada de aquellos que reciben los cuidados (9).

En nuestro país ha aumentado el número de estudios que incluyan este tipo de análisis notablemente hasta alcanzar el 16% de las publicaciones realizadas en España durante el periodo 1983-2008

Si como consecuencia de la intervención o la enfermedad existen cuidados informales, la calidad de vida de los cuidadores debe aparecer diferenciada de aquellos que reciben los cuidados

Por otro lado, no sólo se está incluyendo la medición de la CVRS en estudios de evaluación económica. Otra de las metodologías que vienen incluyendo esta información son los estudios de coste de la enfermedad². En

(1) Health-Related Quality of Life (HRQOL)

(2) Los estudios de costes de la enfermedad (Cost of illness) estiman los costes sanitarios y no sanitarios directos e indirectos que suponen a la sociedad una enfermedad. De este modo identifican el sacrificio social que suponen las enfermedades desde un punto de vista amplio y detallado.

las Islas Canarias se han realizado estudios de coste de la enfermedad en ICTUS, SIDA, Ataxia, Esclerosis lateral, y Alzheimer (10-14). Identificando así no sólo el sacrificio monetario que suponen a la sociedad canaria patologías presentes entre la ciudadanía, sino la percepción de la calidad de vida de los ciudadanos que lo sufren. La incorporación de la CVRS a estos estudios económicos añade una información valiosa

En la percepción de la CVRS entre ciudadanos sanos y enfermos, permite identificar aquellas patologías que mayor pérdida de calidad de vida suponen a los ciudadanos

para los decisores sanitarios poniendo en valor aspectos de valoración subjetiva de los propios usuarios del Sistema Regional de Salud. La continuidad en la recogida de esta información permitirá establecer comparaciones a lo largo de los años en la percepción de la CVRS entre ciudadanos sanos y enfermos, permite identificar aquellas patologías que mayor pérdida de calidad de vida suponen a los ciudadanos y facilita una información valiosa para los investigadores que quieran diseñar estudios más sensibles para identificar los cambios que determinadas intervenciones o políticas sanitarias puedan generar entre la población enferma.

TABLA 1: CALIDAD DE VIDA RELACIONADA CON LA SALUD EN DETERMINADAS PATOLOGÍAS

	SIDA	ICTUS	Depresión	Diabetes	Artritis reumatoide	Alzheimer	Población general
CVRS (EVA)	71,14	53,69	48,43	47,81	47,28	41	59,31

EVA: Escala Visual Analógica

Fuente: Oliva et. al, 2010 (15)

También la CVRS es una variable clave como outcome en estudios que no lleven incluido un análisis de los costes o consumo de recursos. En el 2007 la agencia de evaluación de tecnologías sanitarias de la Comunidad de Madrid utilizaba la calidad de vida relacionada con la salud como una de las variables principales para evaluar las prácticas de cirugía abierta frente a intervenciones laparoscópicas en cirugías aplicadas en el cáncer colorrectal (16). En el mismo

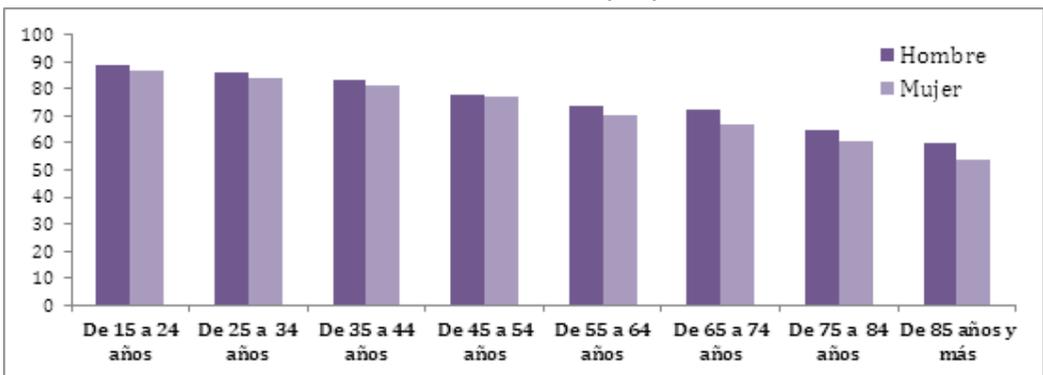
año la agencia catalana de evaluación de tecnologías sanitarias publicó un informe sobre el efecto de intervenciones dirigidas a la práctica de actividad física en la reducción del Índice de Masa Corporal (IMC) entre jóvenes. Además de demostrar que existe eficacia en estas intervenciones para reducir la masa corporal también se observó cómo mejoró la percepción de la calidad de vida y la autopercepción de salud en los individuos que recibieron la intervención (17).

Recientemente se acaban de hacer públicos los resultados de la última Encuesta Nacional de Salud (ENS) en España (18;19).

Entre las numerosas cuestiones de interés (consumo de medicamentos, estancias hospitalarias, peso, altura, etc.) identificadas en esta encuesta destaca la aplicación de una escala muy utilizada entre la comunidad investigadora: la escala EuroQol 5D5L. Esta conocida escala, es una de las herramientas más utilizadas para el cálculo de los Años de Vida Ajustados por Calidad (AVAC). Su incorporación en esta encuesta pone de manifiesto el interés en conocer la calidad de vida relacionada con la salud, siendo la ENS la muestra oficial de mayor envergadura publicada hasta la fecha en la que esta medida ha sido recogida en nuestro país.

Su incorporación en esta encuesta pone de manifiesto el interés en conocer la calidad de vida relacionada con la salud, siendo la ENS la muestra oficial de mayor envergadura publicada hasta la fecha en la que esta medida ha sido recogida

GRÁFICO 1: CALIDAD DE VIDA RELACIONADA CON LA SALUD (EVA) SEGÚN EDAD SEXO.



EVA: Escala Visual Analógica. Fuente: Encuesta Nacional del Salud, 2012.

Hasta ahora en la ENS se preguntaba por el estado de salud con una respuesta cerrada en la que el entrevistado podría responder sobre su estado de su salud entre bueno, malo y regular. En la tabla 2 se muestran el porcentaje de españoles que valoraban positivamente su estado de salud a lo largo de las oleadas llevadas a cabo por el Centro de Investigaciones Sociológicas y el Instituto Nacional de Estadística.

Tabla 2: PORCENTAJE DE PERSONAS QUE VALORAN POSITIVAMENTE SU ESTADO DE SALUD EN ESPAÑA, SEGÚN EL SEXO.

	1987	1993	1995	1997	2001	2003	2006
Ambos sexos	73,89	73,02	71,76	73,16	72,93	71,31	69,98
Varones	77,91	76,63	75,41	76,47	77,96	75,68	75,11
Mujeres	70,04	69,50	68,26	69,92	68,15	67,08	64,97

Fuente: Ministerio de Sanidad, Política Social e Igualdad. Encuesta Nacional de Salud de España (20).

Sin embargo desde el 2004 el Instituto Canario de Estadística y el Servicio Canario de Salud (SCS) vienen incluyendo la misma escala en la Encuesta de Salud en Canarias. De este modo el SCS viene identificando desde el 2004 no sólo la valoración positiva de la salud de los ciudadanos canarios sino también las tarifas resultantes de la aplicación de la escala EuroQol y la valoración de 0-100 (Escala Visual Analógica) sobre su estado de salud (Tabla 3).

En resumen, la medición de la calidad de vida relacionada con la salud se está convirtiendo, a pesar de sus críticas, en un instrumento más dentro del cuerpo de información que manejan los decisores públicos. Utilizar esta información en el diseño de políticas sanitarias y sociales permitirá planificar éstas con un valor añadido hasta ahora ausente. Sin duda ayudará a crear programas que incorporarán la percepción de la salud de los ciudadanos a las que se dirigen.

Tabla 3: PORCENTAJE DE PERSONAS QUE VALORAN POSITIVAMENTE SU ESTADO DE SALUD EN LAS ISLAS CANARIAS, SEGÚN EL SEXO

	VALORACIÓN POSITIVA		EUROQOL EVA	
	2004	2009	2004	2009
Ambos Sexos	69,4	72,85	61,27	74,42
Hombres	74,9	78,36	65,69	77,1
Mujeres	63,9	67,35	56,87	71,78

EVA: Escala Visual Analógica. Fuente: Instituto Canario de Estadística y Servicio Canario de la Salud (21).

Bibliografía

1. Catala-Lopez F, García-Altés A. Evaluación económica de intervenciones sanitarias en España durante el periodo 1983-2008. Rev Esp Salud Pública. 2010;84(4):353-69.
2. Barkun A, Leontiadis G. Systematic review of the symptom burden, quality of life impairment and costs associated with peptic ulcer disease. The American journal of medicine. 2010 Apr;123(4):358-66 e2.
3. Guo JJ, Wade TJ, Keller KN. Impact of school-based health centers on students with mental health problems. Public health reports (Washington, DC : 1974). 2008 Nov-Dec;123(6):768-80.
4. Kim SK, Kim SH, Lee CK, Lee HS, Lee SH, Park YB, et al. Effect of fibromyalgia syndrome on the health-related quality of life and economic burden in Korea. Rheumatology (Oxford, England). 2013 Feb;52(2):311-20.
5. Maklin S, Malmivaara A, Linna M, Victorzon M, Koivukangas V, Sintonen H. Cost-utility of bariatric surgery for morbid obesity in Finland. The British journal of surgery. 2011 Oct;98(10):1422-9.
6. Ackerman I, Ademi Z, Osborne R, Liew D. Comparison of Health-Related Quality of Life, Work Status, and Health Care Utilization and Costs According to Hip and Knee Joint Disease Severity: Results From a National Australian Study. Physical therapy. 2013 Apr 4.
7. Kehusmaa S, Autti-Ramo I, Helenius H, Hinkka K, Valaste M, Rissanen P. Factors associated with the utilization and costs of health and social services in frail elderly patients. BMC health services research. 2012;12:204.
8. Shyu YK, Pan CH, Liu WM, Hsueh JY, Hsu CS, Tsai PS. Health-related quality of life and healthcare resource utilization in Taiwanese women with menopausal symptoms: a nation-wide survey. The journal of nursing research : JNR. 2012 Sep;20(3):208-18.
9. López Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, et al. Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. Gaceta Sanitaria. 2010;24:154-70.
10. Lopez-Bastida J, Oliva Moreno J, Worbes Cerezo M, Perestelo Perez L, Serrano-Aguilar P, Monton-Alvarez F. Social and economic costs and health-related quality of life in stroke survivors in the Canary Islands, Spain. BMC health services research. 2012;12:315.
11. Lopez-Bastida J, Oliva-Moreno J, Perestelo-Perez L, Serrano-Aguilar P. The economic costs and health-related quality of life of people with HIV/AIDS in the Canary Islands, Spain. BMC health services research.

- 2009;9:55.
12. Lopez-Bastida J, Perestelo-Perez L, Monton-Alvarez F, Serrano-Aguilar P. Social economic costs and health-related quality of life in patients with degenerative cerebellar ataxia in Spain. *Movement disorders : official journal of the Movement Disorder Society*. 2008 Jan 30;23(2):212-7.
 13. Lopez-Bastida J, Perestelo-Perez L, Monton-Alvarez F, Serrano-Aguilar P, Alfonso-Sanchez JL. Social economic costs and health-related quality of life in patients with amyotrophic lateral sclerosis in Spain. *Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. 2009 Aug;10(4):237-43.
 14. Lopez-Bastida J, Serrano-Aguilar P, Perestelo-Perez L, Oliva-Moreno J. Social-economic costs and quality of life of Alzheimer disease in the Canary Islands, Spain. *Neurology*. 2006 Dec 26;67(12):2186-91.
 15. Oliva-Moreno J, Lopez-Bastida J, Worbes-Cerezo M, Serrano-Aguilar P. Health related quality of life of Canary Island citizens. *BMC public health*. 2010;10:675.
 16. Soriano M, Fernandez N, Blasco J, Haussmann A, Guerra M. Cirugía laparoscópica en el cáncer colorrectal. Calidad de vida, impacto emocional y satisfacción de los pacientes.: Plan de Calidad para el Sistema Nacional de Salud del Ministerio de Sanidad y Política Social. Unidad de Evaluación de Tecnologías Sanitarias, Agencia Laín Entralgo; 2007. Accesible desde: http://www.madrid.org/cs/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadertype=Content-disposition&blobheadername2=cadena&blobheadervalue1=filename%3D2007_02+CirugiaLaparoscopicaCancerColorrectal.pdf&blobheadervalue2=language%3Des%26site%3DPortalSalud&blobkey=id&blobtable=MungoBlobs&blobwhere=1271807293068&ssbinary=true.
 17. Berra S, López L, López-Aguilà S, Audisio Y, Rajmil L. Sobre peso en la población adolescente: calidad de vida, acceso a los servicios sanitarios y efectividad de las intervenciones: Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya. Plan de Calidad para el Sistema Nacional de Salud del Ministerio de Sanidad y Política Social; 2007. Accesible desde: http://www.gencat.cat/salut/depsan/units/aatrm/pdf/sobrepeso_adolescentes_pcsns07_aatrm2009.pdf.
 18. Encuesta Nacional de Salud (acceso a microdatos) [Base de datos en Internet]. Instituto Nacional de Estadística. 2012. Accesible desde: http://www.ine.es/prodyser/micro_ensalud.htm.
 19. Encuesta Nacional de Salud (acceso a los resultados) [Base de datos en Internet]. Instituto Nacional de Estadística. 2012. Accesible desde: <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t15/p419&file=inebase&L=0>.
 20. Aplicación de consulta. Consulta Interactiva del SNS [Base de datos en Internet]. Ministerio de Sanidad Política Social e Igualdad. 2012. Accesible desde: <http://pestadistico.msc.es/PEMSC25/>.
 21. Comparativa de las Encuestas de Salud de Canarias 2004-2009.: Autovaloración del estado de salud [Base de datos en Internet]. Instituto Canario de Estadística (ISTAC) y Servicio Canario de la Salud (SCS). 2010. Accesible desde: <http://www.gobiernodecanarias.org/istac/jaxi-web/menu.do?path=/03021/C00035A/P0012&file=pcaxis&type=pcaxis>



D. Isaac Aranda

Investigador

Seminario de Investigación en Economía y Salud UCLM

Profesor Ayudante en el área Fundamentos de Análisis Económico de la Universidad de Castilla la Mancha. Licenciado en Administración y Dirección de Empresas.

HOJA DE SÍNTESIS PRÁCTICA

A través de la página BiblioPRO cualquier investigador puede localizar con cierta facilidad y alta sensibilidad distintos cuestionarios según el tipo de patología sobre la que esté investigando: <http://www.bibliopro.org/buscador/>. A través de su buscador avanzado podemos localizar distintas herramientas para identificar y cuantificar correctamente resultados en salud. Otro buscador de instrumentos para registrar medidas de resultado en salud es la web de BIPFAES <http://bipfaes.faes.es>. El acceso a estos dos recursos web es gratuito y sólo hay que rellenar un sencillo cuestionario, sin embargo las ventajas del portal BiblioPRO son mayores que las de BIPFAES, permitiendo realizar pagos on-line, guardar un histórico de descargas, etc.

Existen varias escalas validas para estimar la calidad de vida, para distintos grupos etarios e incluso en algunas específicas para el tipo de enfermedad. Un manual de referencia para identificar estas escalas así como ejemplo donde se ha aplicado es el siguiente:

Preedy VR, Watson R. Handbook of Disease Burdens and Quality of Life Measures. Springer Science+Business Media; 2010. <http://www.springer.com/medicine/book/978-0-387-78664-3>

La iniciativa EUROQUOL distribuye a través de su página Web el cuestionario necesario para calcular las tarifas que sirven para ajustar los años de vida ganados y obtener los Años de Vida Ajustados por Calidad (AVACS). <http://www.euroqol.org> . Este cuestionario está disponible en distintos idiomas: <http://www.euroqol.org/eq-5d-products/eq-5d-5l.html> . En la misma página web se facilitan también las distintas combinaciones de estados de salud para los distintos países en los que está validado: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html> . Para aquellos países en los que aún no ha sido validada esta escala, así como para estudios que hayan utilizado la anterior versión (EQ-5D-3L) también está disponible una matriz de conversión para ajustar los resultados de la versión 3L a 5L. (http://www.euroqol.org/fileadmin/user_upload/Documenten/Excel/Crosswalk_5L/EQ-5D-5L_Crosswalk_Value_Sets.xls)

Otra de las escalas más utilizadas en el ámbito de la evaluación de tecnologías sanitarias para revelar la percepción de la calidad de vida es la escala SF-6D (<http://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d>). Esta escala comparte preguntas de la escala SF-36 (<http://www.sf-36.org/wantsf.aspx?id=1>) y de la escala SF-12 (ampliamente utilizada en el ámbito de la investigación clínica). Por lo tanto si se ha recogido la SF-12 o SF-36 pueden utilizarse para calcular los AVACS (<http://www.qualitymetric.com/Portals/0/Uploads/Documents/Public/SF-6D.pdf>). Recientemente ha sido publicado en la revista Value in Health un

artículo donde se aplica un modelo para la aplicación de esta escala en pacientes con cáncer colorrectal:

Wong CKH, Lam CLK, Wan YF, Rowen D. Predicting SF-6D from the European Organization for Treatment and Research of Cancer Quality of Life Questionnaire Scores in Patients with Colorectal Cancer. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2013;16(2):373-84

La Universidad de Tufts (EE.UU) dispone de un buscador de estudios de coste efectividad a través de la siguiente dirección web: <https://research.tufts-nemc.org/cear4/> . A diferencia de los habituales buscadores de evidencia científica éste incluye una variable que le confiere un gran valor añadido. El registro de estudios de coste efectividad identifica no sólo los artículos y trabajos que han llevado acabo este tipo de análisis sino que informa, antes de hacer accesible el documento completo, los valores que han sido utilizados para el cálculo de los pesos que se utilizaron para evaluar la intervenciones comparadas.

CUADERNOS DE EVALUACIÓN SANITARIA

Los Cuadernos de Evaluación Sanitaria son una colección de artículos breves destinados a presentar los aspectos fundamentales de la Evaluación de Tecnologías Sanitarias, especialmente en el campo de la evaluación económica, dirigidos a Farmacéuticos de Hospital y al resto de profesionales sanitarios del Sistema Nacional de Salud. Su carácter, eminentemente práctico y de síntesis se complementa con una formación a medida que desarrolla los contenidos de cada uno de los números.

La estructura de cada CES es la siguiente:

Un artículo editorial encargado a una firma externa, sobre el tema monográfico del CES, que permita centrar el número y dar una visión desde la experiencia de gestión práctica.

Una serie de artículos cortos, con los temas fundamentales de cada CES, elaborados desde el SIES de la UCLM, el IMW y la FGC.

Identificación de experiencias de éxito en cada uno de los campos tratados por los números del CES y su posible aplicación al SNS.

Una hoja resumen con la información práctica, enlaces y puntos clave de cada uno de los CES