

COLABORACIÓN ESPECIAL**LA MEDICIÓN DE LA CALIDAD DE LOS ESTUDIOS
DE EVALUACIÓN ECONÓMICA. UNA PROPUESTA DE 'CHECKLIST'
PARA LA TOMA DE DECISIONES****José María Abellán Perpiñán, Fernando Ignacio Sánchez Martínez y Jorge Eduardo Martínez Pérez**

Facultad de Economía y Empresa. Departamento de Economía Aplicada. Campus de Espinardo. Universidad de Murcia.

RESUMEN

El propósito del artículo es contribuir a mejorar la eficiencia con que se adoptan decisiones de incorporación de nuevas tecnologías sanitarias, así como orientar la utilización de aquellas por parte de los servicios clínicos. Se sigue un método iterativo para la elaboración de una *checklist* inspirada en los criterios de parsimonia y mensurabilidad. Los autores formularon una versión preliminar sobre la base de la literatura teórica y de las guías de evaluación existentes, que se sometió a diversos tests de validez en el seno de sendos grupos focales compuestos por clínicos y gestores de los servicios regionales de salud de Murcia y Andalucía. El resultado de este proceso iterativo ha sido la confección de una *checklist* que consta de 12 criterios desglosados en varios ítems a los que se asigna una puntuación. Las puntuaciones totales (con un máximo de 100) asignadas a los estudios evaluados por aplicación de la *checklist* se relacionan por medio de una "Guía de uso" con una selección de umbrales coste-AVAC procedente de un trabajo previo de los autores, lo cual permite valorar si la tecnología evaluada es o no coste-efectiva. El instrumento que se presenta, compuesto por la lista de comprobación y una "guía de uso", constituye una combinación inédita en la literatura previa española. A expensas de futuras mejoras en términos de validación y de la extensión de su uso a otros tipos de análisis coste-efectividad, esta propuesta permite formular recomendaciones sobre la financiación y utilización de nuevas tecnologías sanitarias en nuestro país.

Palabras clave: Evaluación. Economía. Coste-efectividad. Técnica delphi.

ABSTRACT**Quality Assessment of Economic
Evaluations in Health Care:
A Checklist and User Guide**

The aim of this paper is to promote the efficiency in the process of incorporating new health technologies, as well as to guide their implementation by physicians. An iterative method has been used to draw a checklist based on parsimony and measurability criteria. Authors made a first version of the checklist on the basis of theoretical literature and economic evaluation guidelines. This preliminary version was discussed and its validity was tested in two focus groups by doctors and managers of the regional public health systems of Murcia and Andalusia. As a result of this iterative process, we present a 12 criteria *checklist* in which a score is assigned to everyone of its items. The overall score a study receives (with a maximum of 100 points) is confronting to a set of cost per QALY thresholds, in order to assess if the technology been evaluated is cost-effective or not. The thresholds was selected from a previous study. We present a checklist and user guide which includes a cost per QALY thresholds matrix. This is an original proposal that has not been previously published in the Spanish literature. Our instrument needs some future improvements in terms of its validation and its spread to other types of cost-effectiveness analysis, apart from those that use QALYs. Nevertheless, our proposal may be useful to provide guidance on the usage and financing of new health technologies in Spain.

Keywords: Economic. Evaluation. Delphi Technique.

Correspondencia:
José María Abellán Perpiñán
Facultad de Economía y Empresa
Departamento de Economía Aplicada.
Campus de Espinardo. Universidad de Murcia
30100 Murcia.
Correo electrónico: dionisos@um.es

Este trabajo ha contado con la financiación de la Consejería de Sanidad de la Comunidad Autónoma de la Región de Murcia en el marco del Convenio de colaboración en materia de Economía de la Salud.

INTRODUCCIÓN

Se ha afirmado que la evaluación económica es el único procedimiento de que se dispone para establecer el valor adicional que aporta una tecnología sanitaria cuando es comparada con una alternativa¹. Cada vez son más los países que parecen reconocer el alcance de esta afirmación, instaurando paulatinamente sistemas centralizados de evaluación económica de nuevas tecnologías médicas (particularmente medicamentos). Una reciente revisión de las experiencias internacionales de este tipo de sistemas (apodados como “de cuarta garantía”) identificaba siete países en los que los fabricantes de nuevas tecnologías sanitarias son requeridos a entregar obligatoriamente información que demuestre que sus productos son coste-efectivos².

El panorama anterior contrasta vivamente con la situación vigente en el Sistema Nacional de Salud (SNS) español. Como dato significativo, baste reparar en que la propia Agencia Estatal de Evaluación de las Políticas Públicas ha señalado recientemente³ la falta de un “procedimiento reglado y basado en la evidencia científica para la calificación del valor terapéutico de los medicamentos a financiar sobre la base de su coste-eficacia como una de las deficiencias del actual Plan Estratégico de Política Farmacéutica para el SNS”.

La situación descrita cambia un tanto cuando se desciende a instancias subcentrales. Son varias las Comunidades Autónomas que cuentan con sus propias agencias de evaluación de tecnologías sanitarias que publican informes en los que se practican evaluaciones económicas. No obstante el impacto efectivo que poseen esas evaluaciones sobre las decisiones sanitarias autonómicas es muy reducido⁴.

Pese a todo lo anterior, hay dos datos que sugieren una progresiva penetración de la evaluación económica en el mundo aca-

démico y clínico. De un lado, el número de estudios de evaluación económica realizados en España no deja de crecer^{5,6}. En segundo lugar, aumenta también el número de comités de evaluación de los centros hospitalarios que tienen en cuenta la información proporcionada en las evaluaciones económicas como criterio adicional para decidir acerca de la incorporación de nuevas prestaciones caracterizadas por su elevado coste/tratamiento (p.ej. citostáticos). El problema en este último caso⁷ radica en que con demasiada frecuencia dichos comités se limitan a comparar el coste de las distintas alternativas, en lugar de efectuar evaluaciones económicas “completas” (comparaciones de costes y resultados).

Ante una realidad como la descrita debería ser bienvenida cualquier herramienta que contribuya a informar sobre criterios propios de la evaluación económica, las decisiones relativas a la incorporación y utilización de nuevas tecnologías sanitarias. Éste es el propósito de los instrumentos conocidos como listas de comprobación o de chequeo (*checklists*).

En el ámbito de la evaluación económica pueden llegar a diferenciarse dos grandes tipos de listas de comprobación. Un primer tipo estaría integrado por aquellas listas que podríamos calificar como “de expertos”. Es decir, listas cuya combinación de dimensiones o secciones en las que se divide la comprobación, e ítems o preguntas que se pretenden responder, obedecen a una selección realizada por investigadores o grupos de expertos en evaluación económica. Dentro de esta categoría las hay de dos clases: las listas que han sido formuladas por uno o varios investigadores a título de propuesta individual, y las que han sido gestadas siguiendo un proceso más o menos formal de “consenso” entre un amplio número de expertos. Entre las primeras tenemos, por ejemplo, la lista propuesta por Drummond et al.¹, de gran popularidad, y recomendada por varias instituciones internacionales⁸, o

la sugerida por Sacristán et al.⁹, que ha sido aplicada en varios estudios^{10,11}. Entre las listas de consenso de expertos, destacamos tres instrumentos, uno más antiguo y los otros dos bastante recientes. El primero correspondería a las recomendaciones formuladas por el Panel de Expertos de EEUU¹², encarnadas en su *reference case* o estudio modelo con el que comparar la calidad de las evaluaciones económicas. Los otros dos instrumentos son, por un lado, la lista conocida como QHES (Quality of Health Economic Studies), promovida por Ofman et al.¹³, y validada por Chiou et al.¹⁴; mientras que la segunda lista ha sido iniciativa de los investigadores afiliados al CHEC-Project (Consensus on Health Economics Criterialist)¹⁵.

El segundo grupo de listas de comprobación que puede encontrarse en el panorama internacional está constituido por aquellos instrumentos que se basan en las recomendaciones incluidas en las guías o directrices de evaluación económica de diferentes instituciones (*guidelines*). Así, hay agencias nacionales que tienen sus propias listas de comprobación (p.ej. NICE en Inglaterra y Gales, o SIGN en Escocia) y, además, han sido confeccionadas bien para poder contrastar el grado de adherencia de los fabricantes a las guías¹⁶⁻¹⁸ bien para examinar la calidad de las evaluaciones publicadas en un país concreto^{19,20}. Por su interés, a los ejemplos anteriores habría que añadir el listado de comprobación que a modo de resumen final incorporan López et al.²¹ a su propuesta de guía para la evaluación económica de tecnologías sanitarias, por tratarse de una aportación acaecida muy recientemente en el territorio nacional.

El objetivo del trabajo es elaborar una lista de comprobación (*checklist*) que permita a los clínicos verificar tanto la relevancia de la evidencia contenida en los estudios de evaluación económica (esto es, que la evaluación en cuestión sirva para averiguar realmente si la tecnología considerada es

coste-efectiva), como su calidad metodológica o validez interna (la coherencia interna del estudio revisado), con el propósito de contribuir a mejorar la eficiencia con que se adoptan decisiones de compra de nuevas tecnologías sanitarias por parte, sobre todo, de los comités de evaluación económica de los centros sanitarios, así como orientar la utilización de aquéllas por parte de los servicios clínicos correspondientes.

PROCEDIMIENTO PARA LA ELABORACIÓN DE LA *CHECKLIST*

Dos son los criterios básicos que desde su primera versión han inspirado la elaboración de la *checklist* que aquí se presenta: a) parsimonia o simplicidad; y b) mensurabilidad o traducción a una puntuación numérica de las respuestas otorgadas a los ítems. El primero de los criterios persigue que el número de aspectos e ítems cubiertos por la guía no fuera excesivo, de modo que estuviese dentro de un rango que no la hiciese inabordable. En este sentido la lista es homologable a instrumentos como el QHES o incluso la lista de Drummond et al.¹

El segundo criterio tenía un doble objetivo. Primero, acrecentar la comparabilidad de los resultados derivados de la aplicación de la lista de comprobación. En segundo lugar, facilitar la formulación de recomendaciones por parte de los usuarios de la lista, a partir de las puntuaciones alcanzadas. En el primero de los aspectos mencionados nuestra lista se asemeja al QHES, ya que este instrumento pondera numéricamente cada uno de los 16 ítems de que consta, al igual que lo hace nuestra lista. En nuestro caso, además, las puntuaciones numéricas se relacionan por medio de una "Guía de uso" elaborada al efecto, con una selección de umbrales coste-efectividad (en concreto, coste por Año de Vida Ajustado por la Calidad), lo cual permite calificar la tecnología evaluada como coste-efectiva,

siempre y cuando su ratio incremental coste-AVAC sea inferior al umbral que le corresponde. Una metodología como la aplicada, que combina una lista de chequeo “cuantitativa” con una matriz con diferentes umbrales de eficiencia, ha sido sugerida en Australia²² como cauce para incorporar en las guías de práctica clínica la evidencia económica sobre tratamientos para el infarto agudo de miocardio.

La lista de comprobación que describimos con detalle en la siguiente sección ha sido elaborada siguiendo un procedimiento iterativo. Quiere esto decir que se partió de una primera versión formulada por los autores a partir de las recomendaciones de las principales guías de evaluación económica², así como atendiendo a las prescripciones normativas de los manuales de referencia en esta área^{1,23}. Esta primera versión fue objeto de discusión en el seno de un grupo focal formado por gerentes, jefes de servicio y de comisiones de evaluación de la red hospitalaria pública de la Comunidad Autónoma de la Región de Murcia para ir depurando el instrumento a través de las sesiones mantenidas. Partiendo de un borrador inicial, a lo largo de dichas sesiones los participantes aplicaron la lista a varios estudios publicados, permitiendo ajustar paulatinamente la selección y redacción de los ítems que componían originalmente el instrumento, constituyendo el resultado de este proceso una primera prueba de validación de su contenido (*content validity*).

Una versión revisada de la *checklist* fue sometida a una segunda ronda de escrutinio en el marco de un seminario de formación en el que participaron médicos del Servicio Andaluz de Salud. En el seno de este grupo se aplicó la lista a los mismos estudios analizados por el primero de los grupos y, adicionalmente, se pidió que uno de los estudios puntuado con la lista de comprobación fuese también evaluado por los participantes con la ayuda de un instrumento de valoración diferente, una

escala visual analógica (figura 1), a modo de test de validez convergente.

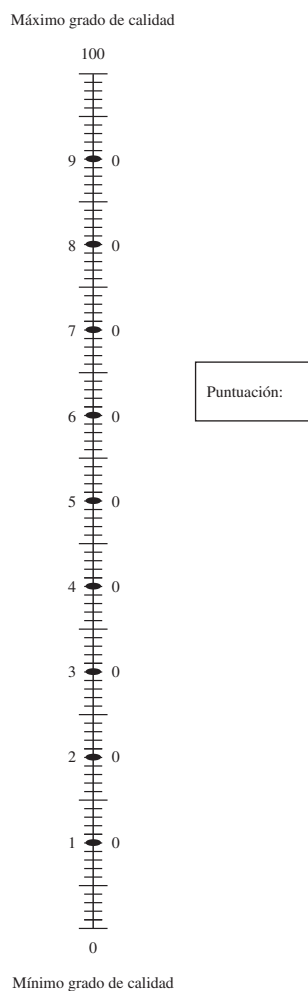
Figura 1

Escala visual analógica utilizada para validar la lista de comprobación

Para ayudarle a manifestar su opinión sobre la calidad metodológica y la validez de los resultados presentados en el artículo titulado.....

.....
hemos dibujado una escala similar a un termómetro. Se considera que el 100 es la puntuación correspondiente al máximo grado de calidad imaginable, y el 0 es la puntuación relativa al mínimo grado de calidad imaginable.

Por favor, en una escala 0-100 como la representada en el termómetro, ¿qué puntuación le daría al citado artículo? Escriba la puntuación en el espacio reservado para ello en el cuadro que figura a la derecha de la escala.



Los resultados alcanzados en las dos pruebas de validación del instrumento fueron satisfactorios. En el primer caso, la impresión subjetiva de los participantes acerca de la claridad y relevancia de los ítems fue buena, a tenor de sus propias manifestaciones. El segundo de los ejercicios de validación, dado su cariz cuantitativo, permitió la realización de algunas pruebas estadísticas. En concreto mediante pruebas no paramétricas se compararon entre sí las puntuaciones conseguidas por el artículo mediante la aplicación de la lista y los valores que cada participante le asignó en la escala visual analógica. No se detectó en ningún caso diferencia estadísticamente significativa entre ambas puntuaciones (prueba Wilcoxon-Mann-Whitney, $p=0,624$). El coeficiente de correlación calculado permite, además, comprobar la magnitud (*effect size*) de la asociación entre las dos puntuaciones (ρ de Spearman = 0,766, $p=0,002$). De acuerdo a las escalas habituales, cabe calificar dicha correlación como "muy fuerte"²⁴ o "alta"²⁵.

SELECCIÓN DE LOS UMBRALES COSTE-AVAC

Los umbrales seleccionados provienen de un trabajo previo²⁶ en el que se identificaron para España los umbrales coste-AVAC basados en la disposición social a pagar por el AVAC (precio sombra del AVAC). Se ha optado por escoger tres "puntos de corte" correspondientes a otros tantos valores de la ratio coste-AVAC. Dado que existen para España dos intervalos estimados de disposición a pagar por un AVAC^{27,28}, se eligió el mayor de los extremos superiores de dichos intervalos (40.200 euros del año 2007) y el menor valor de sus extremos inferiores (8.400 euros), como umbrales máximo y mínimo. El tercero de los puntos de corte se calculó como la media de los extremos seleccionados (24.300 euros).

Puesto que, como más adelante se explica, la aplicación de uno u otro umbral dependerá de la puntuación asignada al estudio en función de su adecuación a los criterios de la *checklist*, resulta de interés saber en qué medida los umbrales seleccionados son o no restrictivos. En particular, conviene conocer hasta qué punto es selectivo el valor más reducido del umbral, pues éste será el que se aplique a aquellos estudios que se alejen en mayor medida del "caso de referencia" definido con la *checklist*, a pesar de haber superado una puntuación mínima.

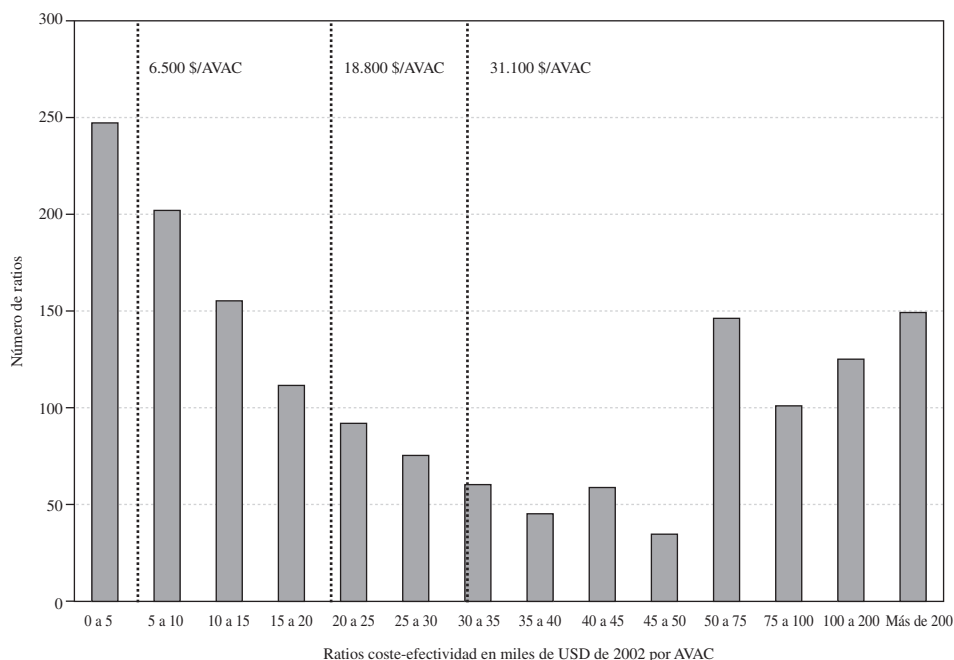
Con este objetivo llevamos a cabo un ejercicio de simulación sobre la base de datos del CEA Registry²⁹, que incluye más de dos mil ratios coste-utilidad procedentes de evaluaciones económicas publicadas, pretendiendo con ello ilustrar la virtualidad de los umbrales seleccionados en la tarea de informar la toma de decisiones. El resultado de aplicar los umbrales seleccionados a la mencionada base de datos se muestra en la figura 2, en la que los umbrales se han convertido a dólares de 2002. De establecer el umbral inferior como "precio sombra" del AVAC, más del 80% de las tecnologías consideradas en la simulación no resultarían coste-efectivas, con un intervalo de confianza al 95% de 78,8%-82,5% (intervalo calculado por *bootstrapping* con 5.000 repeticiones). En el supuesto de aplicar el umbral superior, el porcentaje de tecnologías no coste-efectivas se reduciría hasta, aproximadamente, un 44% (intervalo de confianza al 95%: 41,7%-46,3%). Así pues, al menos con la fuente de información manejada, parece inferirse que los umbrales escogidos son conservadores, pues un porcentaje sustancial de ratios permanecen por encima, incluso, del umbral más laxo de los tres propuestos.

LISTA DE COMPROBACIÓN (CHECKLIST) Y GUÍA DE USO

Los ítems elegidos para formar parte de la *checklist* (tabla 1) se han agrupado en un

Figura 2

Umbral coste/AVAC y ratios incrementales coste-efectividad para una muestra de 1597 estudios de evaluación económica publicados



Fuente: Elaboración propia a partir de CEA Registry²⁹.

total de 12 variables o dimensiones, para lo cual se han tenido en cuenta los posibles usos de la herramienta mencionados en la introducción. Se han omitido algunos de los aspectos que habitualmente son objeto de atención por parte de las guías de evaluación económica, en particular el impacto presupuestario, no porque no sea un tema de singular importancia sino porque su análisis debería ser complementario a la valoración de la solvencia metodológica de la evaluación económica considerada, así como a la idoneidad de la ratio incremental resultante. No en vano algunos autores³⁰ han llegado a calificar el análisis de impacto presupuestario como una “quinta garantía” para los sistemas sanitarios.

Así mismo, puesto que el objetivo del instrumento que proponemos no es el de

evaluar la calidad de los estudios de los que proceden los datos clínicos sobre efectividad, sino la calidad de los estudios de evaluación económica, la lista no hace referencia alguna a esta cuestión, asumiéndose que el análisis crítico de los ensayos o estudios que constituyen la fuente de los datos habrá tenido o tendrá lugar, cuando resulte pertinente, con la ayuda de alguno de los diversos instrumentos existentes al efecto.

Cada uno de los doce criterios o características objeto de evaluación incluye, al menos, dos ítems u opciones a las que se asigna una determinada puntuación. A continuación exponemos brevemente las razones que explican los juicios implícitos en las puntuaciones asignadas a los diferentes ítems de la tabla 1.

Tabla 1

Lista de comprobación (*checklist*) para evaluar la calidad de los estudios de evaluación económica

1. PERSPECTIVA	
1.1a) <i>el estudio adopta una perspectiva social (solamente o en paralelo a la perspectiva del financiador)</i>	8
1.1b) <i>el estudio adopta únicamente la perspectiva del financiador</i>	4
(máximo 8 puntos)	
2. FUENTE DE LA QUE PROCEDE LA INFORMACIÓN SOBRE RESULTADOS Y COSTES	
2.1a) <i>la fuente primaria de información son ensayos clínicos aleatorios de corte "pragmático"</i>	4
2.1b) <i>la fuente primaria de información son ensayos clínicos controlados</i>	2
2.1c) <i>la fuente primaria de información son estudios observacionales</i>	2
2.2a) <i>en los ensayos o estudios las tecnologías de interés se comparan directamente entre sí</i>	4
2.2b) <i>en los ensayos o estudios las tecnologías se comparan indirectamente respecto de un comparador común</i>	2
(máximo 8 puntos)	
3. POBLACIÓN OBJETIVO	
3.1) <i>la población objetivo se describe con detalle</i>	4
3.2) <i>se realiza un análisis de subgrupos para analizar la variabilidad de los resultados debida a características dispares de los pacientes</i>	4
(máximo 8 puntos)	
4. TÉRMINO DE COMPARACIÓN	
4.1a) <i>si existe una práctica dominante (práctica vigente comúnmente utilizada) se ha comparado respecto de la misma</i>	4
4.1b) <i>si no existe una práctica dominante se han efectuado comparaciones múltiples o una combinación lineal de las tecnologías alternativas</i>	4
4.2) <i>se ha comparado respecto de la opción "no hacer nada" o, alternativamente, respecto de la "intervención mínima"</i>	4
(máximo 8 puntos)	
5. MEDIDA DE RESULTADOS	
5.1a) <i>se ha realizado un análisis coste-efectividad con medidas de resultados finales (p. ej. años de vida ganados)</i>	4
5.1b) <i>se utiliza la disposición a pagar como medida de resultados (análisis coste-beneficio)</i>	8
5.1c) <i>se utilizan AVAC como medida de resultados (análisis coste-utilidad)</i>	4
5.2) <i>en caso de utilizar AVAC, los pesos de calidad de vida se han obtenido mediante la lotería estándar o mediante el intercambio de tiempos</i>	4
5.3a) <i>en caso de emplear AVAC, las utilidades se han obtenido a partir de preferencias de la población general, sobre estados de salud definidos por los pacientes</i>	4
5.3b) <i>en caso de emplear AVAC, las utilidades se han obtenido directamente de los pacientes</i>	2
(máximo 12 puntos)	
6. COSTES INCLUIDOS	
6.1) <i>se incluyen todos los costes relevantes dada la perspectiva asumida en el estudio</i>	4
6.2) <i>se ofrece una medida pormenorizada y precisa de los recursos consumidos</i>	4
(máximo 8 puntos)	
7. HORIZONTE TEMPORAL	
7.1) <i>se adopta un horizonte temporal, el mismo para costes y beneficios, lo suficientemente amplio como para tener en cuenta todas las consecuencias asociadas a las tecnologías comparadas</i>	4
7.2a) <i>el estudio dispone de datos primarios que abarquen la totalidad del horizonte temporal</i>	4
7.2b) <i>si sólo se cuenta con datos primarios para el corto plazo, se utiliza un modelo de decisión para extrapolar los datos al largo plazo</i>	4
(máximo 8 puntos)	
8. TASA DE DESCUENTO	
8.1) <i>se aplica una tasa de descuento comprendida entre el 3% y el 5% (inclusive) común a costes y beneficios</i>	4
8.2) <i>se presentan los resultados para tasas de descuento alternativas</i>	4
(máximo 8 puntos)	
9. TRATAMIENTO DE LA INCERTIDUMBRE	
9.1a) <i>se practica un análisis de sensibilidad probabilístico (p.ej. bootstrapping, Montecarlo, ...)</i>	4
9.1b) <i>se practica algún tipo de análisis de sensibilidad determinístico</i>	2
9.2) <i>los resultados del análisis de sensibilidad se presentan en detalle (tablas y gráficos)</i>	4
(máximo 8 puntos)	

Tabla 1 (continuación)

Lista de comprobación (*checklist*) para evaluar la calidad de los estudios de evaluación económica

10) MODELOS DE DECISIÓN EMPLEADOS	
10.1) <i>se detallan los supuestos estructurales del modelo</i> (p.ej. duración de un ciclo en un modelo de Markov)	4
10.2) <i>los resultados del modelo se validan de alguna forma</i> (p.ej. por comparación con los obtenidos por otros modelos publicados para la misma enfermedad e intervención)	4
(máximo 8 puntos)	
11. TRANSFERIBILIDAD	
11.1a) <i>el ámbito de procedencia de los datos coincide exactamente con el de aplicación de la tecnología</i>	8
11.1b) <i>el ámbito de aplicación de la tecnología no coincide con el de procedencia de los datos, pero los resultados del estudio (o estudios) se han adaptado de algún modo al contexto de aplicación</i>	4
11.2) <i>los datos se han obtenido a partir de estudios multinacionales o multicentro, entre los que se encuentra algún centro perteneciente al ámbito de aplicación de la tecnología</i>	4
(máximo 8 puntos)	
12. PRESENTACIÓN DE LOS RESULTADOS	
12.1) <i>los costes y efectos se presentan de forma agregada y desagregada</i>	4
12.2) <i>se calculan y presentan los índices de decisión apropiados</i> (ratios incrementales en ACE y ACU, beneficio neto, ratios beneficio/coste y tasas de rendimiento en ACB)	4
(máximo 8 puntos)	
total apartados 1 a 12 (máximo 100 puntos)	

Los ítems identificados con una letra (p. ej. . 4.1a; 4.1b 7.2a; 7.2b) son excluyentes entre sí.

Por lo que atañe a la perspectiva, consideramos que el carácter universal de la prestación sanitaria y su financiación así mismo indiscriminada a través de los presupuestos públicos permiten concluir que la perspectiva social es la más coherente³¹. Adoptar la perspectiva social no sólo asegura que todos los costes y resultados están incorporados al análisis, sino que exige que los primeros se valoren en términos de coste de oportunidad y que la fuente de obtención de preferencias en la medición de los segundos sea la población general. No obstante, puesto que son mayoría las guías de evaluación que recomiendan la perspectiva del financiador² y siendo relevante la perspectiva restringida en determinados contextos de aplicación, contemplamos esta perspectiva restringida como un “segundo óptimo”.

En lo que se refiere a las fuentes de información sobre costes y resultados, las guías de evaluación muestran un notable consenso acerca de los requisitos que debería cumplir una buena evaluación económica, consenso que tratamos de recoger en el correspondiente apartado de nuestra *check-*

klist. Distinguimos, por un lado, el tipo de fuente de la que proceden los datos, atribuyendo un mayor valor a los ensayos aleatorios de corte “pragmático”, toda vez que con ellos se valora a un tiempo la eficacia y la efectividad de la tecnología. En segundo lugar, se priman los estudios en los que las tecnologías se comparan directamente entre sí frente a aquellos en los que la comparación es indirecta.

La adecuada descripción de la población objetivo se considera una cuestión importante, tanto como el complemento que supone la inclusión de análisis de subgrupos. En este aspecto también existe una recomendación casi unánime en las guías de evaluación.

En relación con el comparador utilizado, la mayor parte de las guías sugieren utilizar la tecnología (o el medicamento o la práctica clínica) habitual o vigente como término de comparación. Dado que esta “práctica habitual” puede ser una combinación de varias, esta posibilidad está contemplada en la *checklist*. Adicionalmente, consideramos conveniente la eva-

luación de la tecnología respecto de la opción "no intervención" o "intervención mínima". Esta estrategia resulta especialmente valiosa cuando el *statu quo* no ha sido objeto de evaluación económica previa, pudiendo ello dar lugar a que su utilización como único comparador sesgase los resultados del estudio.

La concreción de los ítems del criterio número 5 ("Medida de resultados") y su valoración llevan implícitos el juicio de que la unidad de resultados preferida son los AVAC, algo que es compartido por casi todas las guías de evaluación económica. Así, un estudio alcanzará la máxima puntuación en esta dimensión si se trata de un análisis coste-utilidad y, además, los pesos de calidad de vida se han obtenido a partir de una muestra de la población general (perspectiva social), recurriendo a alguno de los métodos de medición de utilidades de estados de salud generalmente admitidos: la lotería estándar (*standard gamble*) o el intercambio temporal (*time trade-off*). Algo inferior será la valoración de una investigación que recurra a la medida monetaria de los resultados, siempre y cuando ésta se base en la disposición a pagar. Finalmente, siendo conscientes de que es práctica extendida el uso de medidas de resultados distintas de los AVAC, se asigna una puntuación inferior a los estudios que miden los resultados en años de vida ganados o unidades similares.

En el apartado de costes, la *checklist* pretende valorar la exhaustividad en la identificación de los costes y el detalle en la medición del uso de recursos.

Por lo que respecta al horizonte temporal, más allá de la exigencia de que éste sea igual para costes y resultados y que su extensión permita recoger todas las consecuencias relevantes, se contemplan dos opciones metodológicas igualmente válidas: que el estudio disponga de datos pri-

marios que cubran todo el horizonte temporal, o bien que se extrapolen los datos mediante el uso de modelos.

A la vista de la evidencia observada en las guías de evaluación revisadas, así como atendiendo a las recomendaciones que se encuentran en la literatura sobre evaluación económica^{1,23}, hemos elegido el intervalo que va del 3% al 5% como referencia para la tasa de descuento, exigiendo que dicha tasa sea idéntica para la actualización de los costes y de los resultados, con el fin de evitar resultados paradójicos³². Se valora, adicionalmente, la presentación de los resultados bajo tasas de descuento alternativas.

En la cuestión relativa al tratamiento de la incertidumbre que afecta a los parámetros y a los modelos, la *checklist* considera separadamente dos cuestiones: la metodología empleada en el análisis de sensibilidad y la presentación detallada, con tablas y gráficos, de los resultados de dicho análisis. Siendo evidente la razón para incluir la segunda de estas cuestiones, tal vez sea preciso explicar que, en relación con la primera cuestión hemos considerado, al igual que las guías de evaluación del NICE y la canadiense, que el análisis probabilístico está por encima del determinístico desde un punto de vista técnico.

Respecto al uso de modelos de decisión, resulta difícil optar por un tipo determinado a la vista de la heterogeneidad observada en la literatura especializada, así como en las propias guías de evaluación. Nuestra *checklist* se limita a valorar el hecho de que se informe adecuadamente acerca de los supuestos del modelo (esto es, la transparencia) y, también, el que se aporte evidencia acerca de la validez del mismo.

La cuestión de la transferibilidad de datos y resultados es otro aspecto que conlleva el consenso de la mayor parte de las guías de evaluación económica existentes,

siendo opinión generalizada que los costes y los consumos de recursos son difícilmente transferibles, mientras que los beneficios sólo lo son en determinadas circunstancias, algo en lo que coincide la literatura³³. Por esta razón, consideramos que el hecho de que el ámbito de procedencia de los datos coincida con el de aplicación de la tecnología evaluada juega a favor de un estudio. En caso contrario, se estima que cierto grado de transferibilidad podrá derivarse de la obtención de datos a partir de estudios multinacionales o multicentro, así como en el caso de que se haya realizado algún tipo de adaptación a un contexto de aplicación diferente.

Por último, el criterio referido a la presentación de los resultados se concreta en dos ítems que recogen, por un lado, la conveniencia de presentar los costes y efectos de forma agregada y desagregada y, por otro, la necesidad de calcular y presentar los índices de decisión propios del tipo de metodología empleada.

Como se ha señalado en la sección anterior, versiones previas de la *checklist* se sometieron a la consideración de sendos grupos de clínicos y gestores sanitarios del Servicio Murciano de Salud y del Servicio Andaluz de Salud, de cuyas apreciaciones se extrajo valiosa información que se incorporó a la versión final que se presenta en la tabla 1. Inicialmente se propusieron dos listas, según que la fuente de procedencia de los datos fuera un único estudio o varios. Tras las rondas con el primer grupo de profesionales se reelaboró una única *checklist*, para lo cual se hubo de modificar el contenido del criterio 2. Otras dimensiones de la lista que sufrieron importantes cambios en su versión final a resultas de los dos procesos de discusión fueron las relativas a la medida de resultados (criterio 5) y transferibilidad (criterio 11). Todas las demás fueron, igualmente, objeto de revisión con el fin de simplificar al máximo su redacción y reducir el número de ítems.

Uno de los objetivos que hemos tenido presente al elaborar esta lista de comprobación ha sido diseñar un instrumento que fuese operativo y de fácil aplicación en distintos contextos, desde la mera revisión crítica de una serie de estudios hasta la toma de decisiones sobre la incorporación de nuevas tecnologías a los centros sanitarios. Esta es la razón por la que, frente el enfoque categórico seguido por otras *checklists* (en las que las respuestas son del tipo “sí/no/no procede”) hayamos optado por puntuar cada ítem, siguiendo el enfoque de instrumentos similares como, por ejemplo, el QHES¹³. Con ello se consigue una mayor operatividad del instrumento a expensas de un cierto (e inevitable) grado de arbitrariedad derivado de la asignación de “números redondos” a las puntuaciones de los distintos criterios, entre los cuales, con la única excepción (poco más que testimonial) de la medida de resultados (criterio 5) se ha optado además por no establecer ningún tipo de ponderación o prioridad. Así, salvo el criterio 5, que permite sumar hasta 12 puntos, todos los restantes permiten sumar como máximo 8 puntos. Se podría decir que en el baremo que se propone se ha primado la sencillez por encima de otras consideraciones.

La forma de aplicar la *checklist* es sencilla: en el caso de que un ítem describa adecuadamente el proceder metodológico del estudio, se asignarán los puntos correspondientes; en caso contrario no se atribuirá puntuación alguna. En todas las dimensiones, la puntuación máxima se obtiene cuando el estudio objeto de valoración coincide con el caso ideal o “caso de referencia”. Una evaluación económica de ese tipo recibiría con este baremo una puntuación total de 100 puntos. Tanto la puntuación máxima como la “guía de uso” de la *checklist* y los umbrales de referencia que a continuación se definen fueron objeto de una profunda revisión a raíz de las opiniones y aportaciones de los clínicos y gestores que participaron en las reuniones de los grupos.

La *checklist* permite, en primer lugar, cuantificar la calidad de la evidencia recogida respecto de la tecnología que nos interesa. Esta utilidad de la lista es independiente del tipo de metodología evaluadora que se hayan utilizado en los estudios revisados: análisis coste-efectividad (ACE), coste-utilidad (ACU) o coste-beneficio (ACB). Con este objeto, creemos adecuado establecer una gradación de la escala de puntos que pueden obtenerse, de forma que habría una puntuación mínima de 40 puntos por debajo de la cual los resultados del estudio no deberían ser tomados en consideración. Así mismo, una puntuación igual o superior a los 60 puntos, sería garantía suficiente para que los resultados de la evaluación se considerasen relevantes. Si la puntuación alcanzada se encontrase entre 40 y 60 puntos, con vendremos que el estudio no es suficientemente informativo, por lo que sus conclusiones deberían ser consideradas con cautela. Es posible imaginar casos en los que la aplicación mecánica de un criterio de exclusión puede no ser la mejor opción. Así, por ejemplo, un estudio calificado con una puntuación comprendida entre 40 y 50 puntos y cuya conclusión fuese que la tecnología evaluada "domina" a la utilizada como comparador (es decir, es más efectiva al tiempo que ahorra costes) constituiría uno de tales casos especiales. Ante una contingencia como ésta la mejor estrategia no sería la de descartar el estudio sino la de tratar de recabar evidencia adicional que decantara el juicio en un sentido u otro.

En segundo lugar, siempre y cuando la metodología de evaluación económica empleada haya sido el ACU, se pueden efectuar recomendaciones acerca de la conveniencia de utilizar/financiar la tecnología considerada, comparando las ratios incrementales coste-utilidad de los estudios con determinados "umbrales" coste/AVAC (*precios sombra* del AVAC). La práctica habitual consiste en recurrir a

"números redondos" (100.000 €/AVAC, 50.000 €/AVAC) o a umbrales basados en "reglas del pulgar" (p.ej. el doble o el triple del PIB per cápita). Frente a este proceder común sugerimos el uso de umbrales basados en preferencias de la población general y obtenidos mediante técnicas de disposición a pagar. Por otra parte, dado que en la literatura existen varios rangos de umbrales coste/AVAC obtenidos de este modo²⁶, la elección del umbral de referencia debería estar relacionada con la solidez metodológica del estudio. Esto es, proponemos aplicar umbrales de referencia cuya magnitud esté directamente relacionada con la puntuación obtenida en el baremo por el estudio de evaluación cuya ratio coste/AVAC se somete a consideración.

El procedimiento sería el siguiente: una vez obtenida la puntuación se acude a la tabla 2 para seleccionar el umbral con el que se debe comparar la ratio coste-efectividad (euros/AVAC) resultante del estudio objeto de evaluación. Si dicha ratio queda por debajo del umbral de referencia la tecnología podría ser objeto de aceptación (es decir, podría considerarse "coste-efectiva"), sin perjuicio de que en función del contexto de utilización de la *checklist* pudiera ser necesaria la valoración de otros aspectos asociados a su adopción o financiación como el impacto presupuestario o los efectos distributivos.

Tabla 2

Umbrales coste/AVAC que deben utilizarse como referencia para recomendar una tecnología evaluada

Puntuación	Umbral (euros/AVAC) [†]
≥ 90	40.200
< 90 pero ≥ 75	24.300
< 75 pero ≥ 60	8.400
< 60 ^{††}	< 0 (ahorro de costes)

[†] El primero de los umbrales se corresponde con el extremo superior del intervalo estimado por Pinto y Martínez²⁶. El segundo coincide con el extremo inferior del intervalo estimado por Pinto y Rodríguez²⁷. El tercero de los valores es el promedio de los dos anteriores. Las cifras se han actualizado a euros de 2007 y redondeado a la centena más próxima.

^{††} Sólo si la puntuación es superior a 40.

CONCLUSIONES

En este trabajo se presenta una lista de comprobación (*checklist*), caracterizada por la parsimonia y la facilidad de uso, concebida para evaluar la calidad de los estudios de evaluación económica. Adicionalmente, se ha desarrollado una guía de uso que contiene una matriz de umbrales coste-AVAC basados en preferencias de los ciudadanos españoles que, en combinación con la *checklist*, permite formular recomendaciones sobre la financiación y utilización de nuevas tecnologías sanitarias. Esta combinación resulta inédita en la literatura española, habiéndose aplicado una aproximación similar únicamente para el caso de Australia²². La lista de comprobación ha sido fruto de un proceso iterativo que ha contado con la participación de sendos grupos de expertos clínicos de la sanidad pública murciana y del sistema sanitario público andaluz. Durante las sesiones mantenidas los participantes valoraron positivamente la versión inicial del instrumento y la guía de uso, y contribuyeron por medio de sus apreciaciones a la manufactura final del mismo.

Aunque la lista de comprobación ha sido objeto de una primera validación en su propio proceso de confección, en lo que respecta a su utilidad para los clínicos, queda por realizar una segunda etapa de validación por un grupo de expertos en evaluación económica. Dicha segunda etapa tendría como objetivo contrastar la similitud entre las puntuaciones generadas por el instrumento y la valoración que emitan dichos expertos. Por ello nos proponemos realizar un estudio de validación semejante al utilizado en el caso del QHES¹³. En consecuencia, la lista de comprobación que se presenta debe entenderse como un instrumento potencialmente perfectible.

Una segunda vía de mejora del instrumento resultaría de ampliar la matriz de ratios coste-AVAC a aquellas situaciones en las que los resultados de las intervenciones

sanitarias se midan en términos de Años de Vida Ganados (AVG). Se trataría por tanto de identificar umbrales coste-AVG basados en preferencias de la población, con los que comparar las ratios incrementales de los estudios publicados. Esperamos de esta forma poder generalizar los posibles beneficios que pudieran derivarse de la aplicación de esta lista de comprobación.

AGRADECIMIENTOS

Los autores agradecen a la Consejería de Sanidad de la Comunidad Autónoma de la Región de Murcia el apoyo financiero recibido en el marco del Convenio en Materia de Economía de la Salud.

BIBLIOGRAFÍA

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ y Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd Edition. Oxford: Oxford University Press; 2005.
2. Del Llano J, Pinto JL y Abellán JM, editores. Eficiencia y Medicamentos: Revisión de las guías de Evaluación Económica. La cuarta garantía. Madrid: Sanofi-Aventis; 2008.
3. Ministerio de Administraciones Públicas. Incidencia de las medidas adoptadas para la racionalización del gasto farmacéutico. Madrid:MAP; 2008.
4. Rodríguez-Monguio R, Antoñanzas F. Health Care Rationing in Spain: Framework, Descriptive Analysis and Consequences. *Pharmaco Economics*. 2006;24:537-548.
5. García-Altés A. Twenty Years of Health Care Economic Analysis in Spain: Are we Doing Well? *Health Econ* 2001;10: 715-719.
6. Oliva Moreno J, del Llano Señarís JE y Sacristán del Castillo JA. La evaluación económica de tecnologías sanitarias en España: situación actual y utilidad como guía en la asignación de recursos sanitarios. *Información Comercial Española* 2003; 804:155-167.
7. Antoñanzas Villar FJ. Challenges to achieving value in drug spending in a decentralized country: the spanish case. *Value Health* 2003; 6 Supl 1:52-63.

8. Kristensen FB, Sigmund H, editors. *Health Technology Assessment Handbook*. Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health; 2007.
9. Sacristán JA, Soto J y Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist, *Ann Pharmacother* 1993; 27: 1.126-1.133.
10. Bradley CA, Iskedjian M, Lanctot KL, Mittmann N, Simone C, St Pierre E, et al. Quality assessment of economic evaluations in selected pharmacy, medical, and health economics journals. *Ann. Pharmacother* 1995;29 (Pt 7): 681-689.
11. Iskedjian M, Trakas K, Bradley CA, Addis A, Lanctôt K, Kruk D, et al. Quality assessment of economic evaluations published in *Pharmacoeconomics*: the first four years (1992 to 1995). *Pharmacoeconomics* 1997; 12:685-94.
12. Weinstein MC, Siegel JE, Gold MR, Kamlet MS y Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996;276:1253-1258.
13. Ofman JJ, Sullivan SD, Neumann PJ, Chiou C, Henning JM, Wade SW y Hay JW. Examining the value and quality of health economic analyses: implications of utilizing the QHES. *J Manag Care Pharm* 2003; 9:53-61.
14. Chiou C, Hay J, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003;41:32-44.
15. Evers S, Goossens M, de Vet H, van Tulder M y Ament A. Criteria List for Assessment of Methodological Quality of Economic Evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005; 21(Pt 2) :240-245.
16. Anis AH, Gagnon V. Using economic evaluations to make formulary coverage decisions. So much for guidelines. *Pharmacoeconomics* 2000;18(Pt1): 55-62.
17. Hill SR, Mitchell AS y Henry DA. Problems with the interpretation of pharmacoeconomic analyses: a review of submissions to the Australian Pharmaceutical Benefits Scheme. *JAMA* 2000;283:2116-2121
18. Ramsberg J, Odeberg S, Engström A y-Douglas L. Examining the quality of health economic analyses submitted to the Pharmaceutical Benefits Board in Sweden. The first year. *Eur J Health Econ* 2004;49:351-356.
19. Atthobari J, Bos JM, Boersma C, Brouwers JRBJ, Jong-van den Berg LTW y Postma MJ. Adherence of Pharmacoeconomic Studies to National Guidelines in the Netherlands. *Pharm World Sci* 2005; 27: 364-370
20. Baladi JF, Menon D y Otten N. Use of economic evaluation guidelines: 2 years' experience in Canada. *Health Econ* 1998;7:221-227
21. López J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, Puig J. Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. Madrid: Plan Nacional para el SNS del MSC. Servicio de Evaluación del Servicio Canario de Salud; 2008. Informes de Evaluación de Tecnologías Sanitarias: SESCS; 2006/22.
22. Antioch KM, Jennings G, Botti M, Chapman R y Wulfsohn V. Integrating cost-effectiveness evidence into clinical practice guidelines in Australia for acute myocardial infarction. *Eur J Health Econom* 2002; 3:26-39
23. Gold MR, Siegel JE, Russell LB y Weinstein MC, editors. *Cost-effectiveness in Health and Medicine*. Oxford: Oxford University Press; 1996.
24. Davis JA. *Elementary survey analysis*. Englewood Cliffs, NJ: Prentice-Hall; 1971.
25. Hinkle DE, Wiersma W, Jurs SG. *Applied statistics for the behavioral sciences*. Chicago: Rand McNally College Publishing.
26. Abellán Perpiñán JM, Sánchez Martínez FI y Martínez Pérez JE. Evaluación económica de tecnologías sanitarias: ¿valen lo que cuestan? *Cuadernos Económicos de ICE* 2008; 75: 191-210.
27. Pinto JL, Rodríguez R. ¿Cuánto vale la pena gastarse para ganar un año de vida ajustado por la calidad? Un estudio empírico. En: *El valor monetario de la salud*. Barcelona: Springer-Verlag; 2001. p.52-91.
28. Pinto JL, Martínez JE. Estimación del valor monetario de los años de vida ajustados por la calidad: estimaciones preliminares. *Ekonomiaz* 2005; 60: 192-209.
29. Center on the Evaluation of Value and Risk in Health. *The Cost-Effectiveness Analysis Registry* [Internet]. (Boston), Tufts-New England Medical Center, ICRHPS. Disponible en: <http://www.tufts-nemc.org/cearegistry/>.
30. Paul JE, Trueman P. 'Fourth hurdle reviews', NICE, and database applications. *Pharmacoeconomics and drug safety*, 2001; 10: 429-438.

31. Brouwer WBF y Koopmanschap MA. On the economic foundations of CEA. Ladies and gentleman, take your positions! *Journal of Health Economics* 2000; 19: 439-459.
32. Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. *Management Science* 1983; 29: 300-6.
33. Welte R, Feenstra F, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. *PharmacoEconomics* 2004; 22:857-876.