

DIFERENCIAS EN LOS VALORES DE CALIDAD DE VIDA RELACIONADA CON LA SALUD OBTENIDOS CON LAS “TARIFAS” SF-6D DE DISTINTOS PAÍSES

Autores: *Jorge Eduardo Martínez Pérez*¹

Fernando Ignacio Sánchez Martínez

José María Abellán Perpiñán

Departamento de Economía Aplicada. Universidad de Murcia

Resumen

La medición de los resultados en salud de las intervenciones sanitarias es una de las tareas críticas de los estudios de evaluación económica. Una de las opciones metodológicas más comunes consiste en obtener las medidas de calidad de vida relacionada con la salud a partir de algoritmos o ‘tarifas’ que asignan valores

¹jorgemp@um.es.

Agradecimientos: Los autores quieren reconocer las valiosas aportaciones realizadas por dos revisores anónimos que, sin duda, han contribuido a mejorar este artículo. Asimismo, desean agradecer el apoyo financiero recibido del Ministerio de Economía y Competitividad a través de los proyectos ECO 2013-48631P y ECO2013-47312P y de la Fundación Séneca mediante el proyecto 19322/PI/14.

a los estados de salud, descritos según un instrumento de medida. En este artículo evaluamos las diferencias existentes entre las tarifas de uno de estos instrumentos de medida, el SF-6D, de España y otros cinco países, y las discrepancias que resultan al aplicarlas a una muestra de población. En ambos casos se encuentran diferencias significativas entre países, siendo la tarifa española la que da lugar a un rango mayor de valores.

Palabras clave: evaluación económica; calidad de vida relacionada con la salud; utilidades; SF-6D; SF-36; tarifas.

Differences in health related quality of life valuations derived from country-specific SF-6D ‘tariffs’

Abstract

Measuring health outcomes is one of the key tasks in economic evaluation studies. One of the most common methodological options consists in deriving health-related quality of life measures from algorithms or ‘tariffs’ which assign values to health states, defined according to a descriptive system. In this paper we assess the differences between tariffs of one of these instruments, SF-6D, in Spain and other five countries, as well as the discrepancies that arise when tariffs are applied to a population sample. In both cases statistically significant differences emerge between countries and the widest rank of values results from the Spanish algorithm.

Key words: economic evaluation; health-related quality of life; health utilities; SF-6D; SF-36, tariffs.

Recibido: 06/09/2016

Aceptado: 24/11/2016

1. INTRODUCCIÓN

Un ámbito de creciente expansión dentro de la economía de la salud es el de la evaluación económica de tecnologías sanitarias. Bajo la denominación genérica de evaluación económica se incluyen diferentes técnicas, como el análisis coste-efectividad (ACE), el análisis coste-utilidad (ACU) y el análisis coste-beneficio (ACB), todas ellas orientadas a contribuir a la asignación eficiente de los recursos en el sistema sanitario. El denominador común de estas técnicas es la

comparación de los costes y los resultados de los programas, tratamientos, intervenciones, etc., con el fin de maximizar las ganancias en salud que se obtienen a partir de un determinado presupuesto (o, alternativamente, minimizar el coste que conlleva lograr un determinado resultado en salud).

La incorporación de la evaluación económica a la toma de decisiones en el Sistema Nacional de Salud (SNS) viene siendo un objetivo declarado por parte de los poderes públicos. No obstante, a diferencia de otros sistemas públicos de salud como los del Reino Unido, Canadá, Australia o Suecia, en España no ha existido hasta la fecha la obligación de someter a un estudio previo de evaluación económica las nuevas tecnologías antes de su entrada al sistema sanitario, lo que se conoce como sistemas de “cuarta garantía”. Un reciente paso en esta dirección han sido la creación, en virtud de un acuerdo de marzo de 2012 del Consejo Interterritorial del SNS, de la Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del SNS, constituida por los órganos evaluadores preexistentes en algunas comunidades autónomas. En este mismo sentido se orienta la exigencia plasmada en el Real Decreto-Ley 16/2012 de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud y mejorar la calidad y seguridad de sus prestaciones (BOE de 24 de abril), de que las nuevas tecnologías sanitarias que aspiren a la financiación pública sean previamente evaluadas por parte de la citada Red.

Más allá de la voluntad política de que esta disposición normativa desemboque en un auténtico sistema de “cuarta garantía” en nuestro país, resulta necesario dotar a las agencias evaluadoras de un marco metodológico común para llevar a cabo sus funciones. Son varias las guías metodológicas de evaluación económica que se han propuesto desde el ámbito académico (Pinto y Sánchez, 2003; López-Bastida et al., 2010; Abellán et al., 2009), sin que quepa identificar importantes discrepancias entre ellas. Recientemente, el servicio regional de salud de Cataluña publicó su guía de evaluación económica y de análisis de impacto presupuestario de medicamentos (CatSalut, 2014).

No son pocos los problemas de índole metodológica que conlleva la tarea de identificar, medir y valorar los recursos empleados (el cálculo de los costes), tarea que se ve, además, constreñida por las limitaciones derivadas de la falta de información o del carácter incompleto de ésta. Sin embargo, donde se plantean los mayores retos es en la medida y valoración de los beneficios de los programas sanitarios, esto es, en la determinación de las ganancias en salud que se derivan de los tratamientos e intervenciones que son objeto de evaluación.

Las técnicas de evaluación económica citadas más arriba (ACE, ACU y ACB), se diferencian, justamente en la unidad de medida de resultados empleada: unidades naturales o clínicas en el primer caso, unidades valorativas o subjetivas en el segundo, y unidades monetarias en el tercero. La ventaja de las dos últimas técnicas citadas es que permiten comparar el grado de eficiencia en el uso

de recursos asociado a cualesquiera tecnologías o intervenciones, por dispares que éstas sean, toda vez que la unidad de resultados es común: años de vida ajustados por la calidad (AVAC) en el caso del ACU y dinero, en el caso del ACB. Por el contrario, al ser la medida de resultados del ACE unidimensional y específica (por ejemplo, saturación de oxígeno en sangre, carga viral, ángulo de giro, etc.), la comparabilidad entre evaluaciones correspondientes a diferentes ámbitos clínicos resulta imposible.

Por otra parte, la dificultad que conlleva la monetización de la salud, imprescindible para traducir los resultados de las intervenciones sanitarias a dinero, como requiere el ACB, es causa de que el uso de esta técnica sea poco frecuente. En consecuencia, el ACU, o lo que es lo mismo, la evaluación económica que utiliza como unidad de resultados los AVAC, se ha convertido en la metodología de referencia y es, de hecho, la opción metodológica recomendada por algunas agencias de evaluación como el National Institute for Health and Care Excellence (NICE) británico.

Los AVAC son una unidad de medida de resultados que combina tiempo de vida con calidad de vida en un único valor. Para su cálculo se requiere medir la calidad de vida relacionada con la salud (CVRS) asociada a cada condición o estado de salud que resulte relevante para la evaluación de una intervención. Aunque existen diversas aproximaciones a la medida de la CVRS, la que resulta útil a los efectos de la evaluación económica es aquella que se basa en la expresión de las preferencias de los individuos (pacientes y/o población general). Son varios los instrumentos que se han desarrollado a tal fin, algunos de los cuales han permitido disponer de conjuntos de valores o “tarifas” de utilidades de estados de salud que pueden ser empleados en los ejercicios de evaluación económica. No obstante, por ser estas tarifas el resultado de estudios de medición de preferencias de ámbito nacional, cabe la posibilidad de que estas preferencias no sean las mismas en distintos países, lo que restaría validez a su utilización en un ámbito geográfico diferente de aquél en el que fueron generados los valores.

En este artículo nos centramos en uno de estos instrumentos de medición de la CVRS en concreto, el SF-6D (Brazier et al., 1998, 2002), con el objetivo de evaluar las diferencias existentes entre las tarifas o conjuntos de valores de utilidad de los países en los que están disponibles, entre los que se encuentra España. El apartado siguiente se dedica a revisar las cuestiones más relevantes referidas a la medición de la CVRS, en general, y a la descripción del instrumento SF-6D, en particular. En el apartado 3 se detalla la metodología seguida, se describen las “tarifas” nacionales que se han utilizado en el trabajo y se ofrecen detalles de la muestra de población a la que se han aplicado los distintos algoritmos. El epígrafe 4 contiene los resultados del análisis comparado de las tarifas y de los valores obtenidos al aplicarlas al perfil de salud de una muestra poblacional. Se cierra el artículo con un apartado de conclusiones.

2. FUNDAMENTOS Y ANTECEDENTES

Idealmente, la medición de la CVRS debería alcanzar todas y cada una de las dimensiones que afectan a la salud de los individuos, abarcando, por tanto, aspectos no solo relativos al estado físico, sino también a la condición psíquica de la personas y a su capacidad para el trabajo, las actividades cotidianas y las relaciones sociales. El interés y preocupación por alcanzar algún tipo de medida “objetiva” de la salud tiene sus orígenes en la Antigua Grecia, si bien los primeros antecedentes en el campo de la literatura económica se remontan a mediados del siglo XX (Karnofsky y Burchenal, 1949). No fue, sin embargo, hasta dos décadas más tarde cuando los instrumentos de medida alcanzaron un cierto refinamiento, siendo su diversidad compatible con la existencia de tres rasgos comunes (Badía y Lizán, 2003):

- a) Su carácter multidimensional, identificándose tres dominios de notable relevancia: físico, psíquico y social.
- b) La naturaleza subjetiva de la evaluación, puesto que cada individuo vive y valora de forma diferente su propio estado de salud.
- c) La expresión cuantitativa (numérica) de las valoraciones –bien sea de modo explícito o implícito–, que, además, se agregan para obtener puntuaciones parciales y/o globales.

A la hora de abordar la medición de la CVRS existen dos enfoques diferentes. De un lado, se encuentra el enfoque psicométrico que, mediante escalas ordinales, persigue aproximar el cambio en la salud experimentado por los pacientes a resultas de un tratamiento o clasificar a estos atendiendo a su nivel de gravedad. De otro lado, la aproximación propia de la Economía de la Salud, esto es, las medidas basadas en preferencias, que emplean valoraciones cardinales que aspiran a cuantificar la intensidad con la que un individuo prefiere un estado de salud frente a otro. Dentro de las medidas psicométricas, se suele distinguir entre medidas específicas y medidas genéricas. Las primeras están diseñadas para ser utilizadas con una determinada patología o tipo de pacientes, pudiendo existir diferentes instrumentos para una misma enfermedad. Sirva como ejemplo la veintena larga de cuestionarios de CVRS existentes para el caso de la diabetes (Sánchez-Lora et al., 2010). Las medidas genéricas, por el contrario, persiguen caracterizar la CVRS de la población general, aunque también pueden ser utilizadas con pacientes de cualquier patología. Algunos ejemplos de este tipo de medidas son el *Sickness Impact Profile* (Bergner et al., 1976), el *Nottingham Health Profile* (Hunt et al., 1985) o el *Short-Form 36 (SF-36)*, propuesto por Ware y Sherbourne (1992). Este último instrumento, el SF-36, es uno de los más empleados (Garratt et al., 2002) y consta de 36 preguntas, agrupadas en 8 dimensiones (función física, rol

físico, dolor corporal, salud general, vitalidad, función social, rol emocional y salud mental). Cada una de estas dimensiones puede tomar un valor comprendido entre 0 y 100 y, además, es posible obtener dos medidas resumen (la componente física y la componente mental). El SF-36, como otros cuestionarios genéricos, cuenta con valores de referencia para la población de distintos territorios (normas poblacionales), por lo que los resultados obtenidos en un estudio para un determinado grupo de población o pacientes pueden ser comparados con dichas normas.

Las medidas que entroncan con el enfoque de la Economía de la Salud, o medidas basadas en preferencias, están constituidas por un sistema descriptivo y un algoritmo que permite generar valores cardinales de intensidad de preferencias, denominados utilidades o pesos de calidad de vida. El sistema descriptivo es similar al de las medidas psicométricas antes citadas, esto es, un conjunto de atributos que intentan capturar la salud del individuo atendiendo a distintas dimensiones, y que puede alcanzar un determinado número de grados o niveles. Entre las más utilizadas cabe citar el *Euroqol-5* dimensiones (EQ-5D), el *Health Utilities Index* (HUI) y el *Short Form-6* dimensiones (SF-6D).

Las medidas basadas en preferencias presentan ciertas ventajas frente a las medidas psicométricas. De un lado, la mayoría de las medidas psicométricas no permiten calcular una puntuación única global por individuo, lo que hace imposible computar un único valor resumen para una población, por agregación de los valores individuales. Además, se ha de tener en cuenta que las medidas psicométricas presentan problemas de agregación, puesto que el modo en que se determinan, en su caso, las puntuaciones resumen conlleva un cierto grado de arbitrariedad. Al no estar basadas en preferencias, las puntuaciones se derivan de criterios ad hoc como, por ejemplo, el consistente en suponer que la distancia entre tener algo de dolor y no tener dolor es la misma que la que existe entre experimentar bastante dolor y mucho dolor. Esto hace que cualquier agregación esté sujeta a controversia (Brazier et al., 1999).

Por el contrario, las utilidades que se obtienen a partir de las medidas basadas en preferencias son válidas para ser utilizadas como pesos o ponderaciones de calidad de vida que, combinados con datos de supervivencia (i.e. cantidad o tiempo de vida), permiten el cálculo de AVAC. Las utilidades se expresan en una escala cuyo valor 0 se corresponde con la muerte y el valor uno con la salud perfecta, siendo posible que determinados estados de salud se consideren más indeseables que la muerte, con lo que su utilidad tendrá signo negativo. La agregación de los años de vida ponderados por la utilidad posibilita el cómputo de la cantidad de salud asociada a un determinado perfil lo que, a su vez, resulta útil para el cálculo de los beneficios en salud que genera un tratamiento o intervención sanitaria (y, por ende, la comparación de la efectividad de diferentes programas) o de la esperanza de vida ajustada por la calidad de una población (Martínez et al., 2016).

Los diferentes instrumentos de medida de la CVRS basados en preferencias tienen, a su vez, ventajas e inconvenientes, derivados de su mayor o menor sensibilidad a la hora de captar determinados problemas de salud o cambios en el estado de salud de los individuos (Brazier et al., 2004; Wee et al., 2007). Precisamente su mayor sensibilidad frente a otros instrumentos genéricos de muy extendido uso como el EQ-5D, es una de las razones que nos ha llevado a centrar el presente estudio en el instrumento SF-6D.

El SF-6D fue propuesto por Brazier et al. (1998) con la intención de poder obtener utilidades a partir del cuestionario de salud SF-36, uno de los instrumentos psicométricos más utilizados en el ámbito clínico. En la versión primigenia del SF-6D cada una de las 6 dimensiones o atributos podía adoptar entre 2 y 6 niveles de gravedad. La segunda y definitiva versión (Brazier et al., 2002) mantuvo las mismas seis dimensiones pero modificó el número de niveles en algunas de ellas, hallándose dicho número en la versión actual entre 4 y 6, según el atributo de que se trate (ver Tabla 1). Así, un estado de salud SF-6D puede describirse con un código numérico que identifica el nivel de gravedad asociado a cada dimensión. Por ejemplo, el estado 111111 se asocia con una situación de perfecta salud, mientras que el estado 645655 describe la peor condición de salud posible. La combinación de las 6 dimensiones con sus correspondientes niveles permite caracterizar un total de 18.000 estados de salud distintos, lo que aporta una gran riqueza informativa.

El segundo elemento que define el SF-6D es el algoritmo de cálculo de utilidades, que no es sino una fórmula polinómica cuyos coeficientes determinan la desutilidad asociada a cada nivel de gravedad de un determinado atributo. Este algoritmo se obtiene a partir de las preferencias de una muestra de la población general sobre un subconjunto de los estados de salud posibles definidos por el sistema descriptivo. Conviene que el subconjunto de estados de salud cumpla con el principio de ortogonalidad, esto es, que contenga el número mínimo de estados de salud distintos que aseguren la posterior estimación, mediante técnicas de regresión, de un modelo aditivo con un grado razonable de validez predictiva.

A la hora de obtener las preferencias sobre los estados de salud de los entrevistados, a partir de las cuales se han estimado posteriormente los algoritmos, dos han sido básicamente los procedimientos utilizados. El primero de ellos se conoce como lotería estándar (*standard gamble*), un procedimiento consistente en pedir al entrevistado que determine el riesgo máximo de muerte que asumiría en un tratamiento hipotético que le permitiría solucionar un problema de salud (descrito como el estado de salud que se pretende valorar). El modo concreto en que se aplicó este procedimiento en el contexto de la obtención del algoritmo del SF-6D en el Reino Unido (Brazier et al., 2002) se expone a continuación.

Tabla 1. El sistema de clasificación de estados de salud SF-6D (SF-36)

Nivel	Funcionamiento físico	Nivel	Dolor
1	Su salud no le limita para realizar <i>esfuerzos intensos</i> .	1	No tiene dolor.
2	Su salud le limita un poco para realizar <i>esfuerzos intensos</i> .	2	Tiene dolor pero no interfiere con su trabajo habitual (fuera de casa o en las tareas del hogar).
3	Su salud le limita un poco para realizar <i>esfuerzos moderados</i> .	3	Tiene dolor que interfiere <i>un poco</i> con su trabajo habitual (fuera de casa o en las tareas del hogar).
4	Su salud le limita mucho para realizar <i>esfuerzos moderados</i> .	4	Tiene dolor que interfiere <i>moderadamente</i> con su trabajo habitual (fuera de casa o en las tareas del hogar).
5	Su salud le limita <i>un poco para bañarse o vestirse</i> por sí mismo.	5	Tiene dolor que interfiere <i>bastante</i> con su trabajo habitual (fuera de casa o en las tareas del hogar).
6	Su salud le limita <i>mucho para bañarse o vestirse</i> por sí mismo.	6	Tiene dolor que interfiere <i>mucho</i> con su trabajo habitual (fuera de casa o en las tareas del hogar).
Nivel	Limitaciones de rol	Nivel	Salud mental
1	No tiene problemas con su trabajo u otras actividades cotidianas a causa de su salud física o de sus problemas emocionales.	1	<i>Nunca</i> se siente muy nervioso o desanimado y deprimido.
2	Ha dejado de hacer algunas tareas en su trabajo o en sus actividades cotidianas a causa de su salud física.	2	Se siente muy nervioso o desanimado y deprimido <i>sólo alguna vez</i> .
3	Hace menos de lo que quisiera hacer a causa de sus problemas emocionales.	3	Se siente muy nervioso o desanimado y deprimido <i>algunas veces</i> .
4	Ha dejado de hacer algunas tareas en su trabajo o en sus actividades cotidianas a causa de su salud física y hace menos de lo que quisiera hacer a causa de sus problemas emocionales.	4	Se siente muy nervioso o desanimado y deprimido <i>casi siempre</i> .
		5	Se siente muy nervioso o desanimado y deprimido <i>siempre</i> .
Nivel	Funcionamiento social	Nivel	Vitalidad
1	Su salud no le dificulta sus actividades sociales <i>en ningún momento</i> .	1	Tiene mucha energía <i>siempre</i> .
2	Su salud le dificulta sus actividades sociales <i>sólo alguna vez</i> .	2	Tiene mucha energía <i>casi siempre</i> .
3	Su salud le dificulta sus actividades sociales <i>algunas veces</i> .	3	Tiene mucha energía <i>algunas veces</i> .
4	Su salud le dificulta sus actividades sociales <i>casi siempre</i> .	4	Tiene mucha energía <i>sólo alguna vez</i> .
5	Su salud le dificulta sus actividades sociales <i>siempre</i> .	5	<i>Nunca</i> tiene mucha energía.

Nota: los ítems del SF-36 que se utilizan para construir el SF-6D son los siguientes: los ítems 1, 2 y 10 de funcionamiento físico, el ítem 3 de limitación del rol debido a problemas físicos, los dos ítems de dolor corporal, los ítems 1 y 4 de salud mental y el ítem 2 de vitalidad. Fuente: Brazier et al. (2002).

En primer lugar, se pide a los entrevistados que elijan entre permanecer en el estado de salud que se pretende valorar y someterse a un tratamiento que puede, bien proporcionarles una salud plena (111111), o bien agravar su situación hasta desembocar en el peor estado de salud que el sistema es capaz de describir (645655), cada una de estas opciones con una cierta probabilidad. Las probabilidades de los dos resultados posibles se van modificando, en función de las respuestas del entrevistado, hasta alcanzar una situación en la que éste sea indiferente entre permanecer en el estado objeto de valoración y someterse al tratamiento propuesto. En segundo lugar, y con el fin de determinar una utilidad normalizada entre el valor 0, correspondiente a la muerte, y el valor 1, representativo de la salud plena, se propone al entrevistado una nueva tarea. En este segundo escenario el estado que se valora es el 645655 y los resultados del tratamiento incierto son la salud perfecta y la muerte. Por consiguiente, la utilidad asociada a cada uno de los estados de salud que componen el subconjunto objeto de valoración se obtiene encadenando las dos respuestas del sujeto.

El segundo de los métodos de elicitación que se ha empleado es una modalidad de lo que se conoce como doble lotería o lotería equivalente (McCord y De Neufville, 1986). Este método persigue, entre otros objetivos, el de superar la limitación que conlleva encadenar loterías, ya que esta forma de proceder suele dar lugar a utilidades “infladas” (Llewellyn-Thomas et al., 1982). En la versión de la doble lotería empleada para la estimación del algoritmo del SF-6D en España (Abellán et al., 2012) el encuestado debía elegir entre dos tratamientos o “loterías”. En la primera lotería los resultados posibles son la salud plena y la muerte, con una distribución de probabilidad que se mantiene invariante. En la segunda lotería los dos posibles resultados son la salud perfecta y el estado que se pretende valorar. Las probabilidades en esta segunda lotería se van modificando hasta que se consigue que el entrevistado se muestre indiferente entre esta lotería y la primera. Puede considerarse que el escenario que plantea la doble lotería se asemeja más a la realidad que el de la lotería estándar, donde el estado que se valora aparece como un resultado seguro, pues en cualquier decisión en el ámbito de la salud siempre existe un determinado grado de incertidumbre.

Una vez obtenidas las valoraciones directas del subconjunto de estados de salud según las preferencias de la muestra de población seleccionada, por cualquiera de los procedimientos descritos, se lleva a cabo un ejercicio econométrico destinado a estimar el algoritmo que permitirá obtener la utilidad asociada a cada uno de los 18.000 estados de salud. También en la fase del análisis de regresión caben dos opciones. En primer lugar, puesto que cada uno de los entrevistados valora varios estados de salud, es posible tratar las respuestas como si se conformasen un panel de datos, siendo la modelización mediante efectos aleatorios la aproximación más común en este caso. También resulta posible estimar modelos de medias por Mínimos Cuadrados Ordinarios (MCO), en cuyo caso, en lugar de tratar las valoraciones individualmente, lo que se pretende explicar es la utilidad media obtenida para cada uno de los estados de salud valorados por los distintos sujetos.

3. MATERIAL Y MÉTODOS

Solo un reducido grupo de países dispone de un algoritmo para el SF-6D, derivado del SF-36. La mayoría de estos algoritmos han seguido, con pequeñas modificaciones, la metodología empleada en el Reino Unido (Brazier et al., 1998; 2002). Es el caso de los estimados en Hong Kong (Lam et al., 2008), Japón (Brazier et al., 2009), Portugal (Ferreira et al., 2010) y Brasil (Cruz et al., 2011). Todos estos estudios utilizan la lotería estándar como técnica de valoración de las utilidades del subconjunto ortogonal de estados de salud y similares procedimientos de estimación econométrica. Las diferencias hay que buscarlas en los tamaños de las respectivas muestras y en el número total de estados de salud valorados directamente. Abellán et al. (2012) utilizaron en el caso de España un método diferente para la valoración directa del subconjunto de estados de salud: una doble lotería, en lugar de la lotería estándar, si bien los modelos de regresión ensayados no difieren de los utilizados en el resto de países.

En relación con esto último, aunque el modelo finalmente escogido no coincide en todos los casos, la mayoría se decanta por uno de los modelos de medias MCO. En el caso de la tarifa británica, además de las variables identificadoras de cada uno de los niveles de las diferentes dimensiones, se recoge una variable de interacción, MOST, que recoge la pérdida adicional de utilidad que se deriva del hecho de que una cualquiera de las seis dimensiones alcance el nivel máximo (el peor).

Los coeficientes que integran los algoritmos de cálculo propuestos en cada país se muestran en la Tabla 2. Dichos coeficientes indican el valor que se ha de restar a la unidad (salud perfecta) en función del nivel en que se encuentre cada una de las dimensiones. Así, por ejemplo, si atendemos al algoritmo estimado para España, un nivel 3 en la dimensión “Funcionamiento Físico” supondría una desutilidad de 0.034. Un nivel 2 en el atributo “Limitaciones de rol” restaría 0.014. De este modo, la utilidad que corresponde al estado 323123 en la tarifa española es 0.705 ($=1-0.034-0.014-0.060-0.000-0.066-0.121$). Con el algoritmo británico se obtendría una utilidad de 0.74 para ese mismo estado.

Como se puede observar, no todos los algoritmos son capaces de discriminar reducciones de utilidad asociadas cada uno de los niveles de las diferentes dimensiones. En ocasiones, dos niveles de una misma dimensión llevan aparejada idéntica reducción en la utilidad (los dos primeros niveles de salud mental en la tarifa británica, por ejemplo). En otras, simplemente, no se deriva pérdida de utilidad alguna por encontrarse un individuo en una de esas situaciones (es el caso, por ejemplo, de los dos primeros niveles de la dimensión dolor para la tarifa portuguesa). Estas “lagunas” en los algoritmos se deben al hecho de que el análisis de regresión arroja como resultado coeficientes no significativos para determinadas variables o inconsistencias entre dichos coeficientes (entendiendo por inconsistencia el que el coeficiente asociado a un determinado nivel de gravedad en un atributo implique una pérdida mayor de utilidad que la que se deriva del coeficiente estimado para un nivel superior).

Tabla 2. Algoritmos para el instrumento SF-6DF estimados en los distintos países (coeficientes de desutilidad de los niveles de cada dimensión) y principales características de los estudios

	España	Reino Unido	Hong Kong	Japón	Portugal	Brasil
Funcionamiento físico						
FF2	0.015	0.035	0.060	0.084	0.029	0.080
FF3	0.034	0.035	0.073	0.057	0.029	0.080
FF4	0.090	0.044	0.099	0.108	0.047	0.080
FF5	0.111	0.056	0.157	0.078	0.050	0.080
FF6	0.338	0.117	0.232	0.144	0.207	0.183
Limitaciones de rol						
LR2	0.014	0.053	0.065	0.062	0.012	0.073
LR3	0.038	0.053	0.053	0.044	0.012	0.073
LR4	0.070	0.053	0.067	0.122	0.061	0.087
Funcionamiento social						
FS2	0.037	0.057	0.052	0.097	0.025	0.000
FS3	0.060	0.059	0.036	0.070	0.025	0.033
FS4	0.203	0.072	0.113	0.088	0.051	0.066
FS5	0.208	0.087	0.133	0.101	0.075	0.078
Dolor						
DO2	0.018	0.042	0.075	0.128	0.000	0.089
DO3	0.034	0.042	0.068	0.069	0.000	0.089
DO4	0.198	0.065	0.082	0.064	0.049	0.089
DO5	0.202	0.102	0.103	0.089	0.049	0.116
DO6	0.318	0.171	0.183	0.115	0.087	0.146
Salud Mental						
SM2	0.066	0.042	0.069	0.099	0.038	0.087
SM3	0.078	0.042	0.037	0.072	0.038	0.087
SM4	0.096	0.100	0.172	0.090	0.066	0.088
SM5	0.224	0.118	0.098	0.084	0.100	0.131
Vitalidad						
VI2	0.058	0.071	0.026	0.077	0.040	0.050
VI3	0.121	0.071	0.031	0.068	0.040	0.050
VI4	0.157	0.071	0.060	0.112	0.041	0.050
VI5	0.199	0.092	0.137	0.088	0.092	0.086
MOST*		0.061				
Muestra (n)	1020	611	128	600	140	494
Estados seleccionados	78	241	49	241	55	248
Método valoración	Doble lotería	Lotería estándar				
Modelo de regresión	Medias (MCO)				Efectos aleatorios	

*Resta utilidad en el caso de que al menos una dimensión se encuentre en el nivel máximo (más grave); si se trata de la dimensión Funcionamiento físico, en los niveles 4, 5 o 6. Fuente: elaboración propia.

Los algoritmos permiten calcular la denominada “tarifa” del instrumento SF-6D en cada país, esto es, el conjunto de las 18.000 utilidades correspondientes a cada uno de los estados que el sistema es capaz de describir. Así, un primer análisis comparativo de las utilidades SF-6D entre países consistirá en evaluar las similitudes y diferencias entre éstos en lo que se refiere a la distribución completa de los valores que integran las respectivas tarifas. Además del análisis gráfico se presentarán los estadísticos descriptivos de cada distribución y se llevarán a cabo pruebas paramétricas (test de igualdad de medias) y no paramétricas (test de Wilcoxon) para contrastar la existencia de diferencias significativas entre las tarifas nacionales.

Habida cuenta de que no todos los estados que describe el SF-6D tienen igual prevalencia en la población (como tampoco entre los pacientes de distintas patologías), cabe la posibilidad de que las diferencias que, en su caso, se pongan de manifiesto al comparar las tarifas resultantes de los algoritmos de cada país no se encuentren (o no en idéntica magnitud) cuando el instrumento se aplica para determinar el estado de salud de una determinada muestra de población. Para indagar en esta cuestión, el análisis de las tarifas se completará con la comparación de los valores de las utilidades que resultan de aplicar los diferentes algoritmos a una misma muestra de población general. A estos efectos se cuenta con una muestra representativa de la población española ($n=2000$) para la cual se dispone, entre otros datos, de la “autoclasiificación” del estado de salud con el instrumento SF-6D. A continuación se describe brevemente la muestra en cuestión.

La Dirección General de Tráfico llevó a cabo un estudio para determinar el Valor Oficial de la Vida Estadística en España (Abellán et al., 2011), y para ello se llevó a cabo una encuesta sobre población adulta no institucionalizada, con muestreo poliétapico estratificado por comunidad autónoma y tamaño de hábitat, donde los municipios y las secciones censales dentro de los mismos se seleccionaron aleatoriamente. Además, se establecieron cuotas por tamaño de hábitat, comunidad autónoma y grupos de edad y sexo, para lograr una muestra representativa de la población española. Como se observa en la Tabla 3, la muestra reproduce las principales características sociodemográficas de la población adulta española, siendo la principal diferencia la que tiene que ver con el nivel de formación alcanzado por la población, pues en la muestra están sobrerrepresentados los individuos con estudios medios, a expensas del resto de niveles formativos.

En la Tabla 4 se recogen los estados de salud más frecuentes entre los individuos de la muestra, siguiendo el esquema descriptivo SF-6D. Como se puede observar, casi un 17% del total de la muestra declara encontrarse en el mejor estado de salud posible según el sistema descriptivo empleado (111111). Si añadimos los siguientes cuatro estados de salud más frecuentes, que suponen un

pequeño alejamiento de la salud plena, se concluye que bastan cinco estados para caracterizar el estado de salud de casi la mitad (49,1%) de la muestra. Pese a ello, se ha de poner de manifiesto que el número total de estados de salud que contiene la muestra de población asciende a 361, de los cuales, 240 corresponden a estados que se observan una única vez en la muestra, 45 a estados que comparten dos observaciones y 23 a situaciones de salud que se repiten únicamente tres veces.

Tabla 3. Caracterización de la muestra de población adulta española y comparación con los valores poblacionales

	Muestra	Población española *
Sexo		
Hombres	49,7	49,3
Mujeres	50,5	50,6
Grupo de edad		
18-24	11,0	9,1
25-44	41,8	39,3
45-64	28,5	30,7
65 y más	18,7	20,8
Nivel de estudios		
Hasta primaria	40,3	54,8
Secundaria	44,9	27,6
Superior	14,8	17,6
Situación laboral		
Empleado	49,1	47,1
Parado	12,4	12,7
Inactivo	38,5	40,1
Estado civil		
Soltero	30,2	31,4
Casado	56,9	56,5
Divorciado	6,5	4,8
Viudo	6,5	7,3

Fuente: elaboración propia a partir de datos del INE. (1) La información referida a la población adulta española se ha obtenido del Censo de Población (2011), para las variables sexo, edad y estado civil, y de la Encuesta de Población Activa (2011) para las variables nivel de estudios y situación laboral.

Tabla 4. Estados de salud presentes en al menos el 0,5% de la muestra de población adulta española según el sistema descriptivo SF-6D

Estado de Salud	n	%	Estado de Salud	n	%
111111	341	16,9	211223	22	1,1
111122	275	13,6	111211	21	1,0
111112	191	9,5	111221	21	1,0
111121	124	6,1	211212	21	1,0
111222	60	3,0	111133	18	0,9
211122	60	3,0	111223	15	0,7
111123	44	2,2	211123	15	0,7
211112	34	1,7	222222	15	0,7
211222	34	1,7	111131	12	0,6
111212	32	1,6	211211	10	0,5
111132	30	1,5	211322	10	0,5
211111	24	1,2	212222	10	0,5
111113	22	1,1	221222	10	0,5

Fuente: elaboración propia.

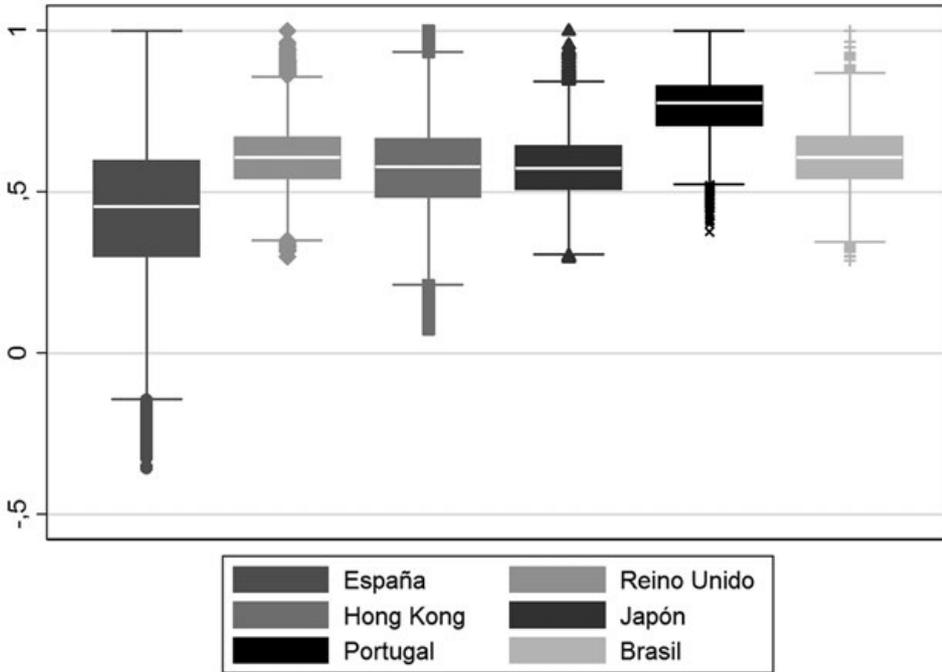
Las distribuciones que resulten de aplicar los algoritmos de la Tabla 2 a los estados de salud declarados por los sujetos que componen la muestra de población serán también objeto de comparación, recurriendo al mismo conjunto de herramientas estadísticas utilizado para comparar las tarifas.

4. RESULTADOS

4.1. Análisis comparado de las tarifas SF-6D

El diagrama de cajas del Gráfico 1 compara las utilidades que se derivan de los diferentes algoritmos de la Tabla 2 o, lo que es lo mismo, los 18.000 valores de utilidad que componen las respectivas tarifas nacionales.

Gráfico 1. Diagramas de cajas de las tarifas nacionales del SF-6D



Fuente: elaboración propia.

A simple vista, las diferencias entre las tarifas resultan evidentes. Los rangos de valores son claramente dispares y se observa como únicamente la tarifa española presenta valores de utilidad negativos. Dicho de otro modo, el algoritmo estimado para España es el único que contempla la posibilidad de que existan estados de salud que se consideren “peores que la muerte”, pues esto es lo que se deduce de la existencia de un valor de utilidad inferior a 0. Por otro lado, también se aprecian diferencias notables en cuanto a la amplitud del recorrido intercuartílico (distancia entre el primer y tercer cuartil). Además, los valores medianos observados son muy distintos, situándose claramente por encima de 0.5 con la única excepción de la tarifa española. Para apreciar mejor la magnitud de estas diferencias, en la Tabla 5 se ofrecen los principales estadísticos descriptivos de cada una de las distribuciones.

Tabla 5. Principales estadísticos descriptivos de las tarifas nacionales del SF-6D

	España	Reino Unido	Hong Kong	Japón	Portugal	Brasil
Mínimo	-0,357	0,301	0,076	0,294	0,378	0,289
Media	0,442	0,607	0,572	0,578	0,761	0,609
Desviación Típica	0,214	0,096	0,131	0,098	0,094	0,098
Error estándar	0,002	0,001	0,001	0,001	0,001	0,001
Percentil 10	0,158	0,484	0,399	0,454	0,627	0,485
Percentil 25	0,301	0,542	0,484	0,508	0,707	0,542
Percentil 50	0,454	0,608	0,577	0,573	0,776	0,607
Percentil 75	0,597	0,670	0,665	0,643	0,829	0,674
Percentil 90	0,712	0,728	0,738	0,707	0,871	0,741
Percentil 95	0,774	0,767	0,781	0,746	0,894	0,777
Percentil 99	0,870	0,844	0,852	0,821	0,934	0,844
Asimetría	-0,300	0,082	-0,179	0,237	-0,633	0,138
Kurtosis	2,830	3,100	2,841	2,908	3,171	2,964

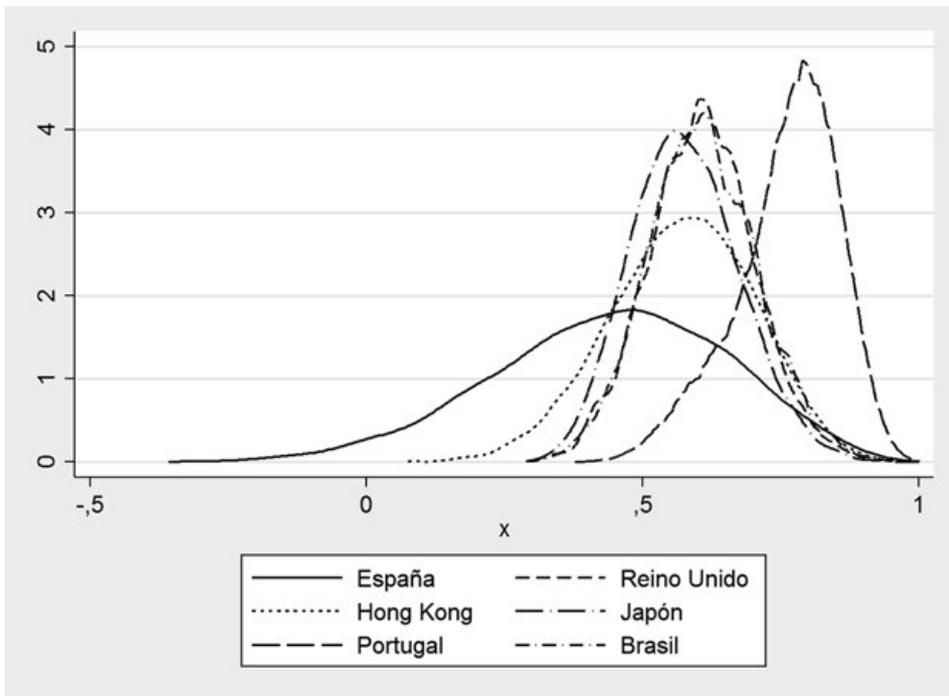
Fuente: elaboración propia.

Por lo que respecta a la media, en la Tabla 5 se aprecia una divergencia notable entre los valores correspondientes a cada país. El valor medio más elevado es el que se deriva del algoritmo portugués (0.761), que supera en más de un 70% el valor medio más bajo, que corresponde a la tarifa española (0.442). También hay diferencias en el grado de dispersión de los valores de las respectivas tarifas, que se puede aproximar por la desviación típica. En efecto, mientras que las utilidades que conforman las tarifas británica, japonesa, portuguesa y brasileña presentan valores de la desviación típica muy similares, en torno a 0.1, la dispersión de las utilidades es significativamente mayor en el caso de Hong Kong (0.13) y muy superior en el de España (0.21). Esta mayor dispersión obedece a que los algoritmos de estos dos últimos países dan lugar a valores mínimos mucho más bajos que el resto. En Reino Unido, Japón y Brasil la utilidad más baja de la tarifa ronda el valor 0.3, en Portugal se acerca 0.4, en Hong Kong el mínimo es inferior a 0.1 y en España, la única tarifa con utilidades negativas, la cota inferior cae por debajo de -0.3.

De los valores registrados en los percentiles que se recogen en la tabla se infiere que la forma de las distribuciones es muy diferente. Además, el coeficiente de kurtosis sugiere una mayor concentración de valores de utilidad alrededor de la media para las cuatro primeras distribuciones señaladas más arriba, siendo el achatamiento de la distribución mucho mayor para el caso de la tarifa hongkonesa, y,

especialmente, para el de la española. En cuanto a la asimetría, algunas tarifas, como la española, la presentan a la izquierda, mientras que otras, como la japonesa, la presentan a la derecha. Estas diferencias distribucionales son apreciables en el Gráfico 2.

Gráfico 2. Distribuciones de las utilidades generadas por los algoritmos nacionales del SF-6D



Fuente: elaboración propia.

Pese a que las diferencias señaladas hasta el momento parecen evidentes, resulta necesario completar este análisis comparativo con la realización de pruebas de carácter paramétrico y no paramétrico que confirmen la discrepancia observada en los valores puntuales de las medias y de las medianas. En la Tabla 6 se constata la existencia de diferencias estadísticamente significativas entre las medias de todas las tarifas, con la única excepción de la comparación entre la tarifa británica y la brasileña ($p\text{-value}=0.101$). El test no paramétrico refuerza la conclusión de que las distribuciones de las utilidades SF-6D difieren significativamente entre países (solo para el caso de Reino Unido vs. Brasil se obtiene un $p\text{-value} > 0.05$).

Tabla 6. Test de igualdad de medias y test de Wilcoxon entre las distintas tarifas nacionales SF-6D (p-value)

Test de Medias	Reino Unido	Hong Kong	Japón	Portugal	Brasil
España	0,000	0,000	0,000	0,000	0,000
Reino Unido		0,000	0,000	0,000	0,101
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
Test de Wilcoxon	Reino Unido	Hong Kong	Japón	Portugal	Brasil
España	0,000	0,000	0,000	0,000	0,000
Reino Unido		0,000	0,000	0,000	0,080
Hong Kong			0,003	0,000	0,000
Japón				0,000	0,000
Portugal					0,000

Fuente: elaboración propia.

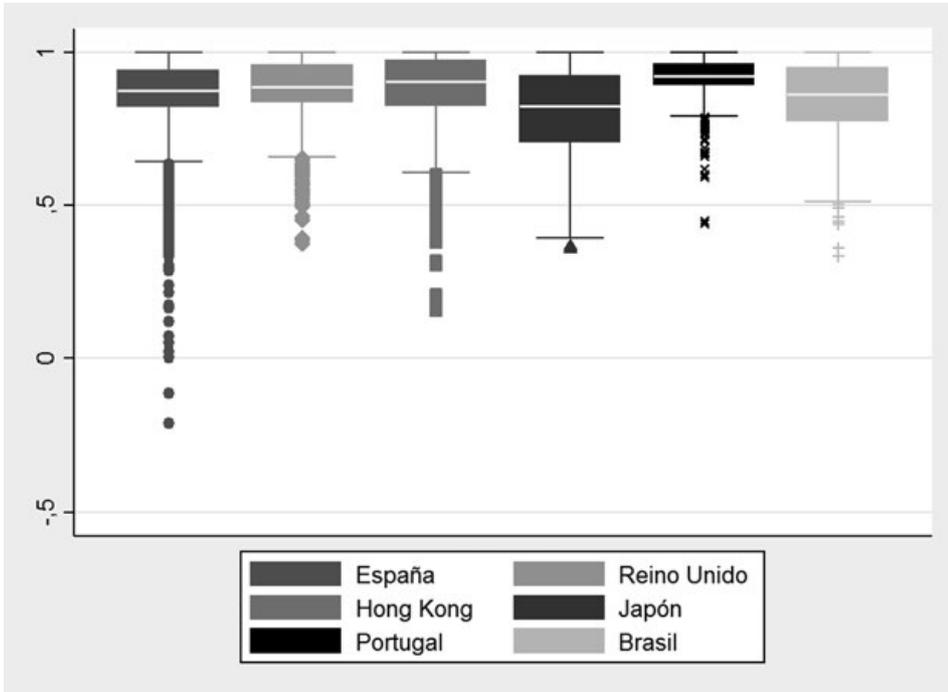
4.2. Diferencias entre algoritmos al ser aplicados a una muestra de población

Como se señaló en la sección anterior, es posible que las discrepancias observadas entre las tarifas nacionales (resultado, a su vez, de las diferencias entre los algoritmos estimados en cada país), no se manifiesten con igual intensidad cuando las tarifas se aplican a muestras de población general o a grupos de pacientes. La razón que subyace a esta hipótesis es la desigual prevalencia de los estados de salud descritos por el instrumento SF-6D. Así, por ejemplo, la mayor amplitud del rango de utilidades que se observa en las tarifas hongkonesa y, sobre todo, española puede perder relevancia en la práctica si en la muestra de población no están representados los estados de salud más graves, que son los responsables de que las distribuciones de las dos tarifas citadas se “alarguen” por su lado izquierdo.

Para analizar el comportamiento de las utilidades obtenidas con las distintas tarifas para una muestra común de población, seguiremos una estructura de análisis similar a la empleada en el epígrafe precedente. En primer lugar, el diagrama de cajas (Gráfico 3) muestra como, al circunscribirse el análisis a un subconjunto de estados de salud, los rangos son en todos los casos sustancialmente menores a los observados en el Gráfico 1, si bien persisten importantes diferencias en función de cuál sea el algoritmo nacional empleado. De igual modo, la mediana de las utilidades se encuentra mucho más próxima a la unidad y el

recorrido intercuartílico se ha reducido de forma más que notable para todas las tarifas, sin excepción.

Gráfico 3. Diagramas de cajas de las utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D



Fuente: elaboración propia.

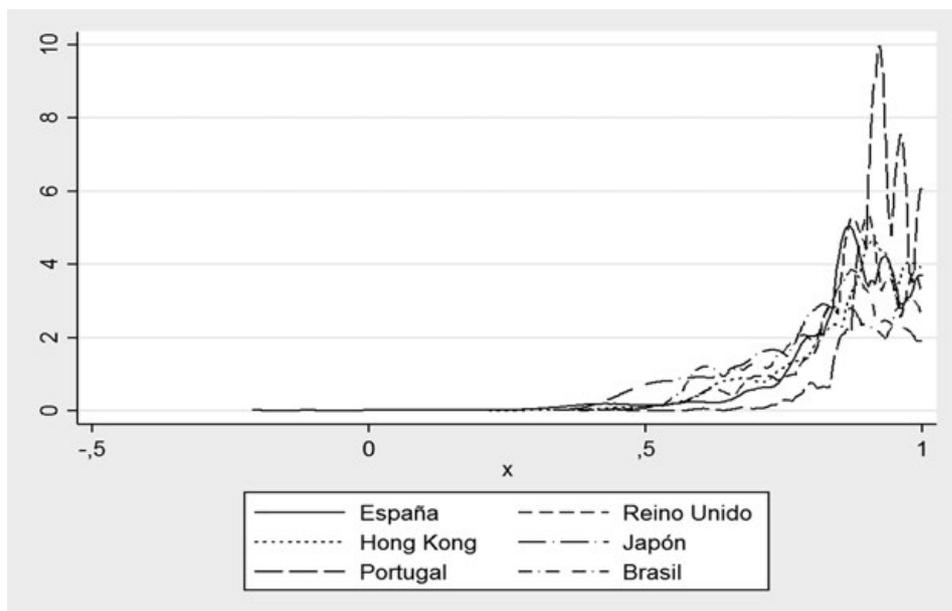
La comparación de los valores puntuales para los principales estadísticos descriptivos queda recogida en la Tabla 7. Vuelve a resultar llamativa la disparidad entre el valor mínimo que se obtiene empleando la tarifa española y el que se deriva de aplicar el resto de algoritmos. No obstante, como consecuencia de la reducción del número de estados de salud que forman las respectivas distribuciones, se registra un estrechamiento de las diferencias en lo que respecta a los valores medios y medianos. Esto se refleja, asimismo, en el estadístico que mide la asimetría de la distribución que, a diferencia de lo observado en la Tabla 5, toma ahora valor negativo para todos los casos. La menor dispersión lleva también aparejado un incremento sustancial de la kurtosis, especialmente en el caso de los valores obtenidos con los algoritmos español y portugués. La mayor similitud entre las distribuciones queda igualmente en evidencia en la representación del Gráfico 4.

Tabla 7. Principales estadísticos descriptivos las utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D

	España	Reino Unido	Hong Kong	Japón	Portugal	Brasil
Mínimo	-0,211	0,377	0,156	0,358	0,440	0,333
Media	0,863	0,873	0,871	0,804	0,928	0,840
Desviación Típica	0,139	0,107	0,121	0,152	0,059	0,126
Error estándar	0,003	0,002	0,003	0,003	0,001	0,003
Percentil 10	0,716	0,700	0,691	0,565	0,856	0,621
Percentil 25	0,823	0,838	0,826	0,705	0,893	0,774
Percentil 50	0,876	0,887	0,905	0,824	0,922	0,863
Percentil 75	0,942	0,958	0,974	0,923	0,962	0,950
Percentil 90	1,000	1,000	1,000	1,000	1,000	1,000
Percentil 95	1,000	1,000	1,000	1,000	1,000	1,000
Percentil 99	1,000	1,000	1,000	1,000	1,000	1,000
Asimetría	-2,377	-1,175	-1,384	-0,603	-1,674	-0,643
Kurtosis	11,649	4,473	5,499	2,573	10,058	2,709

Fuente: elaboración propia.

Gráfico 4. Distribuciones de las utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D



Fuente: elaboración propia.

En cualquier caso, resulta preciso testar si este estrechamiento de las distancias observadas en las medias y en las medianas de las distribuciones se traduce en la desaparición de las diferencias estadísticamente significativas que se constataron al comparar las respectivas tarifas nacionales. El resultado de las pruebas paramétricas y no paramétricas, comparando por pares las distribuciones de las utilidades obtenidas con la muestra de población general, se muestra en la Tabla 8.

Tabla 8. Test de igualdad de medias y test de Wilcoxon entre las distribuciones de utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D (p-value)

Test de Medias	Reino Unido	Hong Kong	Japón	Portugal	Brasil
España	0,010	0,038	0,000	0,000	0,000
Reino Unido		0,657	0,000	0,000	0,101
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
Test de Wilcoxon	Reino Unido	Hong Kong	Japón	Portugal	Brasil
España	0,000	0,025	0,000	0,000	0,000
Reino Unido		0,902	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000

Fuente: elaboración propia.

Por lo que atañe a las medias, parece que las diferencias se mantienen, en general. Solo resulta imposible rechazar la hipótesis de igualdad de medias en las comparaciones de las utilidades calculadas con el algoritmo del Reino Unido y las que resultan de los algoritmos hongkonés (p-value: 0.657) y brasileño (p-value=0.101). En este último caso, se confirma lo que se observó al comparar las utilidades de los 18.000 estados. Si atendemos al resultado del test no paramétrico la conclusión es análoga, dado que únicamente desaparecen las diferencias para el caso de la comparación entre los valores derivados de la tarifa británica y de la hongkonesa, persistiendo en el resto de comparaciones por pares practicada.

El tamaño de la muestra y su representatividad por grupos de edad nos permiten profundizar un poco más en el análisis, con el fin de comprobar si, para los distintos grupos etarios, se confirman las diferencias en los valores medios y medianos de las utilidades que se derivan de los diferentes algoritmos. La Tabla 9 recoge estos valores para los cuatro grupos de edad en lo que hemos segmentado

la muestra, constatándose, en primer lugar, la relación inversa esperada entre edad y utilidad (o CVRS). Parece, a la vista de los datos de la tabla, que las diferencias son, en general, mayores en términos relativos cuanto mayor es la edad. La explicación a este fenómeno se haya en que, según se produce dicho avance, es menor el porcentaje de individuos que afirma disfrutar de un buen estado de salud y, consiguientemente, aparecen mayores diferencias por los distintos valores que cada algoritmo atribuye a cada uno de los niveles de gravedad.

Tabla 9. Medias y medianas de las utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D, por grupos de edad

	España	Reino Unido	Hong Kong	Japón	Portugal	Brasil
Medias						
18 a 24 años	0,925	0,930	0,936	0,892	0,958	0,912
25 a 44 años	0,901	0,908	0,912	0,854	0,945	0,883
45 a 64 años	0,845	0,854	0,850	0,770	0,918	0,814
65 años y más	0,768	0,789	0,773	0,689	0,888	0,743
Medianas						
18 a 24 años	0,934	0,929	0,932	0,901	0,962	0,913
25 a 44 años	0,909	0,894	0,914	0,851	0,959	0,863
45 a 64 años	0,876	0,887	0,887	0,795	0,922	0,861
65 años y más	0,813	0,810	0,775	0,680	0,893	0,710

Fuente: elaboración propia.

Las pruebas de significatividad estadística confirman esta impresión aunque no en todos los casos (Tabla 10). Los tests de medias arrojan un primer resultado de interés: las diferencias entre las utilidades medias obtenidas a partir de la tarifa británica y de la española no resultan estadísticamente significativas (al 5%) en ninguno de los grupos de edad (p-values de 0.415, 0.078, 0.229 y 0.074 para los respectivos grupos en orden ascendente de edad). Lo mismo ocurre cuando la comparación se realiza entre la utilidad media obtenida a partir del algoritmo británico y la derivada del algoritmo hongkonés (p-values 0.397, 0.257, 0.555 y 0.110, respectivamente). Esta ausencia de diferencias estadísticamente significativas entre las medias se observa en alguna otra comparación pero no para la totalidad de grupos de edad. Si atendemos a los resultados de las pruebas no paramétricas, únicamente en el caso de los dos grupos de mayor edad resulta imposible afirmar que existan diferencias significativas entre los valores que surgen de las

tarifas británica y española ($p=0.309$ para el grupo de 45 a 64 años y $p=0.915$ para el grupo de 65 años y más). Esta ausencia de diferencias estadísticamente significativas según el test de Wilcoxon se constata también entre los valores resultantes de los algoritmos de Hong Kong y España para el grupo de 45 a 64 años ($p=0.895$) y entre las utilidades derivadas de la tarifa hongkonesa y la del Reino Unido para el grupo de 25 a 44 años ($p=0.072$).

Tabla 10. Test de igualdad de medias y test de Wilcoxon entre las distribuciones de utilidades obtenidas para una muestra de población general española (n=2000) con los diferentes algoritmos del SF-6D. (p-value), por grupos edad

Test de Medias					
	Reino Unido	Hong Kong	Japón	Portugal	Brasil
18 a 24 años					
España	0,415	0,102	0,002	0,000	0,095
Reino Unido		0,391	0,000	0,000	0,013
Hong Kong			0,000	0,000	0,001
Japón				0,000	0,033
Portugal					0,000
25 a 44 años					
España	0,078	0,005	0,000	0,000	0,000
Reino Unido		0,257	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
45 a 64 años					
España	0,229	0,528	0,000	0,000	0,000
Reino Unido		0,555	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
65 y más años					
España	0,074	0,672	0,000	0,000	0,041
Reino Unido		0,110	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,004
Japón				0,000	0,000
Portugal					0,000

Tabla 10 (cont.)

Test de Wilcoxon					
	Reino Unido	Hong Kong	Japón	Portugal	Brasil
18 a 24 años					
España	0,001	0,011	0,000	0,000	0,000
Reino Unido		0,000	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
25 a 44 años					
España	0,000	0,000	0,000	0,000	0,000
Reino Unido		0,072	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
45 a 64 años					
España	0,309	0,895	0,000	0,000	0,000
Reino Unido		0,000	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000
65 y más años					
España	0,915	0,000	0,000	0,000	0,000
Reino Unido		0,000	0,000	0,000	0,000
Hong Kong			0,000	0,000	0,000
Japón				0,000	0,000
Portugal					0,000

5. CONCLUSIONES

La evaluación económica de tecnologías sanitarias se presenta como una herramienta de apoyo a la toma de decisiones en el sistema público de salud, proporcionando información acerca de la relación coste-efectividad de las nuevas tecnologías (esto es, si las innovaciones terapéuticas, organizativas o de otro tipo “valen lo que cuestan”). Existe un cierto consenso metodológico en torno al modo en que han de realizarse los ejercicios de evaluación económica, que se refleja en las guías de evaluación adoptadas por aquellos países en los que existen

sistemas de “cuarta garantía” y en las propuestas que desde el ámbito académico y desde algunas agencias evaluadoras se han realizado en nuestro país.

Uno de los problemas que se plantean a la hora de llevar a cabo estudios de evaluación económica es la obtención de los inputs (costes y resultados) que constituyen la base de tales estudios. En ocasiones no se dispone de datos primarios procedentes del mismo entorno en el que la tecnología objeto de evaluación se prevé aplicar, lo que obliga a recurrir a fuentes de información indirectas, como bases de datos de costes o medidas de resultados existentes en la literatura, cuya validez podría estar comprometida en la medida en que los contextos en los que estos datos se han generado sean muy distintos de aquellos en los que se realiza la evaluación.

En este artículo nos hemos centrado, en particular, en los efectos que puede suponer aplicar algoritmos o “tarifas” de utilidades generadas en un determinado país a datos sobre el estado de salud de otro, con el fin de generar medidas de resultados en salud susceptibles de ser utilizadas en ejercicios de evaluación económica. Concretamente, hemos evaluado las diferencias que afloran al comparar los distintos algoritmos estimados para el instrumento SF-6D en seis territorios: Reino Unido, Hong Kong, Japón, Portugal, Brasil y España. Adicionalmente, hemos aplicado cada uno de estos algoritmos a los estados de salud declarados por una muestra representativa de la población adulta española.

La primera conclusión que se obtiene del análisis de los algoritmos es que existen diferencias significativas entre las tarifas de prácticamente todos los países. El caso de España resulta especialmente singular, pues el algoritmo SF-6D español es el que da lugar a un rango mayor de valores y el único capaz de generar utilidades negativas para los estados de salud más graves. La explicación de esta discrepancia hay que buscarla, principalmente, en el hecho de que para el algoritmo español se utilizaron medidas directas de utilidad obtenidas mediante un método de doble lotería, distinto al procedimiento de elicitación empleado en el resto de países: la lotería estándar. Existe evidencia de que la lotería estándar genera una aversión al riesgo extrema, lo que da lugar a valores de utilidad más elevados que los obtenidos con otros métodos de elicitación (Wakker y Deneffe, 1996; Bleichrodt et al., 2001; 2007). Este efecto puede verse, además, reforzado cuando el método se aplica a través de un encadenamiento, como es el caso de las tarifas analizadas, con la excepción de la española (Rutten-van Mólken et al., 1995; Bleichrodt, 2001; Oliver, 2003).

Sin embargo, no parece que esta pueda ser la única explicación de las discrepancias entre las tarifas, por cuanto las cinco restantes, muy similares en lo que respecta a la metodología seguida para su obtención, también muestran diferencias estadísticamente significativas en las pruebas paramétricas y no paramétricas. Solo entre las tarifas del Reino Unido y Brasil resulta imposible rechazar la hipótesis de igualdad.

Estudios previos han identificado diferencias significativas en los valores de las utilidades SF-6D medidas directamente en distintos países que sirvieron de base para la posterior estimación de los algoritmos nacionales. Así, Kharroubi (2015)

encontró diferencias “potencialmente importantes” entre las valoraciones de los estados de salud obtenidas mediante la técnica de la lotería estándar en los estudios de Reino Unido y Japón. Estas diferencias cabría atribuir las, según el autor, a diversos factores, que van desde el ámbito lingüístico (la traducción podría no ser del todo equivalente) hasta un posible patrón cultural en la actitud hacia el riesgo (las utilidades son, en general, mayores en Japón, lo que revela una mayor aversión al riesgo). Similares resultados se encuentran en Kharroubi et al. (2014) para el caso de las tarifas de Reino Unido y Hong Kong.

Tan interesante como poner de manifiesto las diferencias de origen cultural o de otro tipo que existen en la medición de las utilidades de estados de salud en distintos países, así como en las tarifas nacionales que a partir de ellas se obtienen, es constatar su relevancia en la práctica. Esto es, hasta qué punto las diferencias en los algoritmos se trasladan a las utilidades que con ellos se generan en una muestra de población. El resultado que hemos obtenido al aplicar los algoritmos a una muestra de población española arroja resultados similares a los que se derivan de comparar las tarifas: las distribuciones de las utilidades que cada algoritmo genera para las descripciones de salud de la muestra presentan diferencias estadísticamente significativas. De nuevo, como ocurría con las tarifas, la distribución resultante de aplicar el algoritmo español es la que presenta un valor mínimo más bajo y una mayor dispersión. Pero también los valores que se deducen del algoritmo hongkonés muestran una elevada dispersión y una cota inferior sensiblemente menor que la de las utilidades obtenidas con el resto de algoritmos.

La principal lección que se deriva de este análisis comparado es que, siempre que ello sea posible, en los estudios de evaluación económica deberían valorarse los resultados con instrumentos de medida de la CVRS basados en preferencias de la misma población a la que va dirigida la intervención que se desea evaluar. Las diferencias resultantes de aplicar tarifas estimadas en otros países pueden llegar a ser de una considerable magnitud, lo que podría comprometer seriamente la validez de los indicadores coste-efectividad obtenidos en el análisis.

Más allá de diferencias metodológicas en la estimación de las “tarifas”, las preferencias en relación con el estado de salud pueden diferir notablemente entre países, de suerte que el valor de las nuevas tecnologías, medido en términos de ganancia en la salud, puede ser bien distinto en función de cuál sea el origen de las preferencias sobre las que tal valor se asienta.

BIBLIOGRAFÍA

- ABELLÁN, J. M. y otros (2009), “La medición de la calidad de los estudios de evaluación económica. Una propuesta de ‘checklist’ para la toma de decisiones”, *Revista Española de Salud Pública*, vol. 83, núm. 1, pp. 71-84.

- ABELLÁN, J. M. y otros (2011), *El valor monetario de una vida estadística en España. Estimación en el contexto de los accidentes de tráfico*, Dirección General de Tráfico, Madrid [disponible en: <http://www.msssi.gob.es/profesionales/saludPublica/prevPromocion/Lesiones/JornadaDecenioAccionSeguridadVial/docs/InformeVVEJorgeMartinez.pdf>; acceso septiembre 2016].
- ABELLÁN, J. M. y otros (2012), “Lowering the ‘floor’ of the SF-6D scoring algorithm using a lottery equivalent method”, *Health Economics*, núm. 21, pp. 1271-1285.
- BADIA, X. y LIZÁN, L. (2003), “Estudios de calidad de vida”, en Martín, A. y Cano, J. (coords), *Atención primaria: conceptos, organización y práctica clínica*, Alianza Editorial, Madrid, pp. 250-261.
- BERGNER, M. y otros (1976), “The sickness impact profile: conceptual formulation and methodological development of a health status measure”, *International Journal of Health Services*, núm. 6, pp. 262-271.
- BLEICHRÖDT, H. (2001), “Probability weighting in choice under risk: an empirical test”, *Journal of Risk and Uncertainty*, núm. 23, pp. 185-198.
- BLEICHRÖDT, H. y otros (2001), “Making descriptive use of prospect theory to improve the prescriptive use of expected utility”, *Management Science*, núm. 47, pp. 1498-1514.
- BLEICHRÖDT, H. y otros (2007), “Resolving inconsistencies in utility measurement under risk: test of generalization of expected utility”, *Management Science*, núm. 53, pp. 469-482.
- BRAZIER, J. y otros (1998), “Deriving a preference-based single index from the UK SF-36 health survey”, *Journal of Clinical Epidemiology*, vol. 51, núm. 11, pp. 1115-1128.
- BRAZIER, J. y otros (1999), “A review of the use of health status measures in economic evaluation”, *Health Technology Assessment*, vol. 3, núm. 9, pp. 1-164.
- BRAZIER, J. y otros (2002), “The estimation of a preference-based measure of health from the SF-36”, *Journal of Health Economics*, núm. 21, pp. 271-292.
- BRAZIER, J. y otros (2004), “A comparison of the EQ-5D and SF-6D across seven patient groups”, *Health Economics*, vol. 13, núm. 9, pp. 873-884.
- BRAZIER, J. y otros (2009), “Estimating a preference-based index from the Japanese SF-36”, *Journal of Clinical Epidemiology*, núm. 62, pp. 1323-1331.
- CATSALUT (2014), *Guía y recomendaciones para la realización y presentación de evaluaciones económicas y análisis de impacto presupuestario de medicamentos en el ámbito del CatSalut*, Servei Català de la Salut, Barcelona.
- CRUZ, L. N. y otros (2011), “Estimating the SF-6D value set for a population-based sample of brazilians”, *Value in Health*, núm. 14, pp. 108-114.

- FERREIRA, L. N. y otros (2010), "A portuguese value set for the SF-6D", *Value in Health*, vol. 13, núm. 5, pp. 624-630.
- GARRAT, A. M. y otros (2002), "Quality of life measurement: bibliographic study of patient assessed health outcome measures", *British Medical Journal*, núm. 324, pp. 1417-1421.
- HUNT, S. y otros (1958), "Measuring health status: a new tool for clinicians and epidemiologists", *Journal of the Royal College of General Practitioners*, núm. 35, pp. 185-188.
- KARNOFSKY, D. A. y BURCHENAL, J. H. (1949), "The clinical evaluation of chemotherapeutic agents in cancer", en MacLeod, C. M. (ed), *Evaluation of Chemotherapeutic Agents*, Columbia University Press, New York, pp. 191-205.
- KHARROUBI, S. A. y otros (2014), "A comparison of Hong Kong and United Kingdom SF-6D health state valuations using a nonparametric bayesian method", *Value in Health*, vol. 17, núm. 4, pp. 397-405.
- KHARROUBI, S. A. (2015), "A comparison of Japan and UK SF-6D health state valuations using a non-parametric bayesian method", *Applied Health Economics and Health Policy*, vol. 13, núm. 4, pp. 409-420.
- LAM, C. L. K. y otros (2008), "Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population", *Value in Health*, vol. 11, núm. 2, pp. 295-303.
- LLEWELLYN-THOMAS, H. y otros (1982), "The measurement of patients' values in medicine", *Medical Care*, núm. 17, pp. 501.
- LOPEZ-BASTIDA, J. y otros (2010), "Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias", *Gaceta Sanitaria*, vol. 24, núm. 2, pp. 154-170.
- MARTINEZ, J. E. y otros (2016), "Esperanza de vida ajustada por la calidad en España: una aproximación", *Revista de Evaluación de Programas y Políticas Públicas*, núm. 6, pp. 28-48.
- McCORD, M. y DE NEUFVILLE, R.(1983), "Empirical demonstration that expected utility decision analysis is not operational", en Stigum, B. P y Wenstop, F. (eds), *Foundations of Utility and Risk Theory with Applications*, D. Reidel, Dordrecht, pp. 181-199.
- OLIVER, A. (2003), "The internal consistency of the standard gamble: tests after adjusting for prospect theory", *Journal of Health Economics*, núm. 22, pp. 659-674.
- PINTO, J. L. Y SÁNCHEZ, F. I. (2003), *Métodos para la evaluación económica de nuevas prestaciones*, Ministerio de Sanidad y Consumo, Madrid.
- RUTTEN-VAN MÖLKEN, M. P. y otros (1995), "Methodological issues of patient measurement. Experience from two clinical trials", *Medical Care*, núm. 33, pp. 922-937.

- SANCHEZ-LORA, F. J. y otros (2010), “Instrumentos específicos de medida de calidad de vida relacionada con la salud en diabetes mellitus tipo 2 disponibles en España”, *Medicina Clínica*, vol. 135, núm. 14, pp. 658-664.
- WAKKER, P. y DENEFFE, D. (1996), “Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown!”, *Management Science*, núm. 42, pp. 1131-1150.
- WARE, J. E. y SHERBOURNE, C. D. (1992), “The MOS-36 item short-form health survey (SF-36): I. Conceptual framework and item selection”, *Medical Care*, núm. 30, pp. 473-483.
- WEE, H. L. y otros (2007), “Assessing differences in utility scores: a comparison of four widely used preference-based instruments”, *Value in Health*, vol. 10, núm. 4, pp. 256-265.