

**UNIVERSIDAD DE
MURCIA**



**FUNCIÓN DE UTILIDAD MULTIATRIBUTO PARA EL INSTRUMENTO DE
MEDIDA DE LA CALIDAD DE VIDA RELACIONADA CON LA SALUD SF-6D^(*)**

Enero 2010

Equipo investigador

José María Abellán Perpiñán

Jorge Eduardo Martínez Pérez

Ildefonso Méndez Martínez

Fernando Ignacio Sánchez Martínez

Grupo de Investigación en Economía de la Salud y Evaluación Económica

Departamento de Economía Aplicada

Universidad de Murcia

^(*) Informe correspondiente al proyecto *Diseño y estimación de una nueva tarifa social SF-6D mediante el enfoque de la teoría de la utilidad multiatributo*, desarrollado en el marco del Convenio de Colaboración suscrito con fecha de diciembre de 2009 entre la C.A.R.M.-Consejería de Sanidad y la Universidad de Murcia, en materia de Economía de la Salud, - BORM N.º, --/01/2010.



ÍNDICE

1. Introducción.....	3
2. Antecedentes.....	7
2.1. Sistemas de clasificación multiatributo para la medición de la calidad de vida relacionada con la salud	7
2.2. La obtención de funciones de utilidad multiatributo a partir del enfoque descompuesto o algebraico.....	12
3. Motivación y objetivos del estudio.....	19
3.1. Motivación.....	19
3.2. Objetivos.....	20
4. Material y métodos.....	21
4.1. Modelización y procedimientos de medición de las utilidades	21
4.1.1. La función de utilidad multiatributo del SF-6D	21
4.1.2. Obtención de las utilidades: método de elicitación	23
4.1.3. Cálculo de las utilidades	26
4.2. Recogida de datos	27
4.2.1. Las muestras	27
4.2.2. Los cuestionarios	28
4.3. Evaluación de la capacidad predictiva de la función estimada	30
5. Resultados.....	33
5.1. Participantes en el estudio	33
5.2. Utilidades de los estados de salud SF-6D y utilidades “uni-atributo”	35
5.2.1. Utilidades multiatributo	35
5.2.2. Utilidades uniatributo.....	36
5.3. Análisis.....	37
5.3.1. Obtención de los parámetros de escalamiento	37
5.3.2. Función de utilidad multiatributo para el SF-6D	39
5.3.3. Comparación de la función multiatributo con el algoritmo SF-6D obtenido mediante inferencia estadística.	46
6. Conclusiones.....	55
6.1. Función de utilidad multiatributo para el SF-6D	55
6.2. Comparación de la función multiatributo con las tarifas estimadas mediante inferencia estadística.....	58
6.3. Aplicaciones de la función de utilidad multiatributo	59
Referencias	61
Apéndice 1. Los cuestionarios.....	65
Apéndice 2. Ejemplo de aplicación de la función de utilidad multiatributo a partir de las respuestas al cuestionario SF-36.	73

1. Introducción

En el marco de los proyectos incluidos en la primera edición del Convenio entre la Consejería de Sanidad de la C.A.R.M. y la Universidad de Murcia en materia de economía de la salud¹, el equipo investigador de la Universidad abordó la estimación de una función de utilidad multiatributo para el instrumento de clasificación de estados de salud conocido como SF-6D, en lo que constituyó la primera estimación de la “tarifa” de utilidades de este instrumento para España. En aquella ocasión se aplicó el enfoque basado en inferencia estadística, por otra parte el único previamente utilizado en la literatura para estimar el algoritmo del SF-6D y el más comúnmente empleado en la obtención de tarifas de utilidades para instrumentos semejantes, como el EQ-5D.

Al margen de otros resultados que se reflejaron en la memoria final justificativa del Convenio, el proyecto aludido tuvo como principal logro la obtención de un algoritmo para el SF-6D, basado en las preferencias de una muestra de población general española (concretamente, de la Región de Murcia). Dicho algoritmo se estimó a partir de las valoraciones de una selección de estados SF-6D con un método de “doble lotería” (*probability lottery equivalence*); un procedimiento de elicitación novedoso en este campo y susceptible de escapar a algunos de los sesgos que parecen afectar a los métodos habituales, como la “lotería estándar” (*standard gamble*). Adicionalmente, se obtuvieron modelizaciones alternativas de la función de utilidad del SF-6D, resultado de aplicar un ajuste semiparamétrico a los modelos de efectos aleatorios convencionales, garantizándose con ello que los posibles sesgos de selección presentes en la muestra no afectasen a las estimaciones de los parámetros del modelo.

El enfoque de inferencia estadística utilizado en el citado estudio constituye una de las dos vías metodológicas que cabe seguir a la hora de derivar tarifas o paneles de utilidad para instrumentos de clasificación de los estados de salud multiatributo, como el SF-6D, el EQ-5D o el HUI (*Health Utilities Index*). La vía alternativa a este enfoque, también denominado “compuesto”, se basa en los postulados la teoría de la utilidad

¹ Convenio de Colaboración suscrito con fecha 22 de diciembre de 2006 entre la C.A.R.M.-Consejería de Sanidad y la Universidad de Murcia, en materia de Economía de la Salud, - BORM Nº 23, 29/01/2007.



multiatributo y recibe el nombre de enfoque “descompuesto” o algebraico. El interés, no solo teórico sino también práctico, por contrastar la validez de nuestras estimaciones paramétricas y semiparamétricas, así como analizar la robustez de los algoritmos obtenidos para el SF-6D respecto del procedimiento utilizado para su obtención, nos llevó a proponer una extensión de nuestro primer estudio consistente en derivar un nuevo algoritmo para el SF-6D recurriendo, en este caso, al enfoque algebraico.

Este proyecto se incluyó en el protocolo de la segunda edición del Convenio en materia de economía de la salud entre la CARM y la Universidad de Murcia², al amparo del cual se desarrolló su primera fase, y cuya conclusión ha tenido lugar a lo largo del año 2009 con la cobertura del Convenio suscrito en dicho año. En la primera fase del proyecto se abordó el diseño de la investigación, se elaboraron los nuevos cuestionarios asistidos por ordenador y se realizó el trabajo de campo. Este último iba orientado a obtener, a partir de una muestra representativa de la población general en la Región de Murcia, un conjunto de valoraciones que, unidas a una parte de las obtenidas en el estudio inicial, permitieran derivar la función de utilidad del SF-6D partiendo de un enfoque “descompuesto”.

La diferencia principal entre este enfoque y el basado en inferencia estadística reside en el hecho de que el enfoque algebraico descompone la valoración de los estados de salud multiatributo en evaluaciones separadas de cada uno de los atributos o dimensiones que permiten describir el estado de salud. Así, las encuestas realizadas en el curso de la primera fase de este proyecto tenían como objetivo construir las funciones de utilidad “uniatributo” (es decir, de cada atributo considerado de manera aislada). Estas funciones de utilidad, junto con las utilidades multiatributo de un reducido número de estados procedentes del primer estudio, son la base para la obtención de la función de utilidad multiatributo del SF-6D.

En el presente informe se da cuenta de los resultados finales de este proyecto que se inició en el año 2008, así como del proceso seguido a lo largo de estos dos años transcurridos hasta su conclusión. Los resultados obtenidos son, por otra parte, objeto de comparación con los que surgieron del primer estudio basado en métodos de

² Convenio de Colaboración en materia de Economía de la Salud suscrito con fecha 27 de noviembre de 2008 entre la C.A.R.M.-Consejería de Sanidad y la Universidad de Murcia. BORM Nº 300, 29/12/2008.



inferencia estadística. Esta comparación reviste un gran interés desde el punto de vista académico, por cuanto sólo en una ocasión hasta la fecha se ha abordado la estimación del algoritmo de un sistema de clasificación de los estados de salud multiatributo a través de ambos enfoques.³ Las estimaciones realizadas por el equipo investigador de la Universidad de Murcia no sólo son pioneras en España, sino que permiten someter a análisis las conclusiones del único estudio existente, habida cuenta de las innovaciones que incorpora nuestro diseño de investigación, particularmente en lo que se refiere al método de elicitación empleado.

Como expresábamos en las primeras páginas de la memoria del proyecto de modelización de la tarifa del SF-6D basada en métodos de inferencia estadística, confiamos en que los resultados de esta investigación, las conclusiones que de ellos se derivan y las propuestas que se sugieren en la parte final de esta memoria sirvan, cuando menos, para acrecentar los argumentos que informen la reflexión, previa a la toma de decisiones, por parte de los gestores públicos y los decisores políticos.

³ McCabe et al. (2005a, b) utilizaron ambas aproximaciones con el instrumento HUI:2. En Stevens et al. (2007) se comparan los resultados.

2. Antecedentes

2.1. *Sistemas de clasificación multiatributo para la medición de la calidad de vida relacionada con la salud*

2.1.1. *Medidas de calidad de vida relacionada con la salud*

La calidad de vida relacionada con la salud (CVRS) es un concepto que idealmente abarcaría todos los atributos o dimensiones que potencialmente puedan caracterizar a la salud, lo que requiere considerar tanto el ámbito de la función física como el plano psicológico, o incluso el rol social. Tras el concepto de CVRS se esconden no obstante dos aproximaciones muy diferentes cuando se afronta la tarea de su medición. En general, es posible seguir dos enfoques provenientes de tradiciones científicas distintas: el enfoque de la Psicología (enfoque *psicométrico*) y el enfoque propio de la Economía de la Salud (o simplemente *económico*). Los dos enfoques tienen en común su fundamento en la *percepción subjetiva* del paciente, pero difieren en sus objetivos y en los medios que emplean. Las medidas o escalas de CVRS del enfoque psicométrico pretenden reflejar el cambio en la salud experimentado por un paciente como efecto del tratamiento, o pretenden discriminar entre pacientes con diferentes niveles de gravedad, utilizando para ello cuestionarios en los que las respuestas se miden en escalas ordinales tipo Likert. En cambio, las *medidas basadas en las preferencias* del enfoque económico –también denominadas escalas de utilidad multiatributo– intentan medir la intensidad con que el paciente prefiere un estado de salud a otro, empleando para tal finalidad escalas cardinales definidas por los puntos 0 y 1 correspondientes a la muerte y la salud plena, respectivamente.

Dentro del primer grupo de medidas (las escalas de CVRS no basadas en las preferencias) resulta habitual distinguir entre medidas *específicas* y medidas *genéricas* (o perfiles de salud). Las primeras pretenden medir la CVRS asociada a pacientes afectados por una patología específica, siendo ejemplos de tales medidas, entre otros, *Ashma Quality of Life Questionnaire*, *Skindex-29* o *Diabetes Quality of Life Questionnaire*. Algunas de las medidas genéricas más importantes son *Nottingham Health Profile*, *Short Form-36*, o *Sickness Impact Profile*. Aunque su sensibilidad al cambio cabe esperar que sea inferior a la correspondiente a las medidas específicas, su mayor ductilidad las convierte en una herramienta muy valiosa cuando lo que interesa



es contar con un criterio homogéneo (para todos los individuos igual) de clasificación del estado de salud.

2.1.2. Escalas multiatributo: el componente descriptivo.

Las medidas de CVRS basadas en preferencias se caracterizan por constar de dos elementos: un sistema descriptivo multiatributo de carácter genérico, y una regla de valoración o algoritmo que sirve para generar los índices cardinales de bienestar (intensidad de preferencias) o utilidades. El primero de estos dos componentes es análogo en su naturaleza a cualquiera de las medidas genéricas de CVRS que hemos citado antes. Las únicas diferencias afloran en cuanto a las dimensiones y niveles que integran, pero no en la estructura general que se utiliza para describir el estado de salud. Por tanto, su rasgo distintivo radica en el mencionado algoritmo que acompaña al componente meramente descriptivo. Algunos de los ejemplos más conocidos de son el *EuroQol* (EQ)-5D, el SF-6D y el *Health Utilities Index* (HUI).

El *Health Utilities Index* tiene su origen en el instrumento desarrollado por un grupo de investigadores de la Universidad de McMaster, Canadá, (Torrance et al., 1982; Boyle et al., 1983) para su uso en la evaluación de los resultados de salud producidos por la asistencia intensiva a neonatos. Dicho sistema, constaba de cuatro dimensiones (función física, función de rol, función socio-emocional y problema de salud) y un rango de niveles entre 4 y 8, con el que se podían llegar a definir hasta 960 estados de salud, y pasó a conocerse con el nombre de *Health Utilities Index Mark 1* (HUI:1). La segunda versión de este instrumento, el HUI:2 (Torrance et al., 1996) incrementó la riqueza informativa del original, con el fin de extender su aplicabilidad al ámbito de los cuidados a niños con cáncer, utilizando para ello siete atributos (sensación, movilidad, emociones, cognición, auto-cuidado, dolor y fertilidad), con un número variable de niveles (3-5), que permitían generar 24000 estados de salud. Finalmente, el HUI:3 (Furlong et al., 1998; Feeny et al., 2002) ofrece un sistema descriptivo aún más completo, compuesto por ocho atributos de 5-6 niveles cada uno: visión (6 niveles), oído (6), habla (5), desplazamiento (6), destreza (6), emociones (5), cognición (6), y dolor (5). El poder descriptivo de este instrumento es tal que puede caracterizar hasta 972000 estados de salud diferentes.

Por su parte, el *Euroqol* 5-D (*EQ-5D*) nació de una revisión sistemática de otras medidas genéricas de CVRS que incluyó el *Quality of Well-Being*, el *Sickness Impact Profile*, el *Nottingham Health Profile* y el índice de Rosser, efectuada por el Grupo EuroQol (EuroQol Group, 1990). La versión inicial del sistema actual constaba de seis dimensiones, con 2 ó 3 niveles por dimensión. Los primeros estudios (Essink-Bot et al., 1990; Kind, 1990; Brooks et al., 1991; Nord, 1991) aconsejaron reducir a cinco las dimensiones (‘actividades de ocio’, cuya contribución a la descripción del estado de salud se reveló marginal, se integró con ‘actividad principal’ en una única dimensión denominada ‘actividades cotidianas’), al tiempo que se optó por fijar tres niveles para todos los atributos –ningún problema, algunos problemas, problemas extremos–. Nació así el sistema actual con cinco dimensiones (5D) –movilidad, cuidado personal, actividades cotidianas, dolor/malestar, ansiedad/depresión– y tres niveles de gravedad por dimensión –ningún problema, algunos problemas y problemas extremos–. El sistema descriptivo EQ-5D puede generar hasta 243 (3^5) estados de salud diferentes.

El SF-6D nace como un intento de intermediar entre una medida genérica de CVRS preexistente (el SF-36 primero y, posteriormente, el SF-12) y el algoritmo generador de utilidades que todo instrumento basado en preferencias posee. La principal motivación que llevó a Brazier et al. (1998) a tender ese “puente”, respondió al interés que suponía contar con una medida basada en preferencias que conservase parte de la riqueza informativa albergada en los cuestionarios *Short Form*. A diferencia del HUI o el EQ-5, el SF-6D siempre ha caracterizado la salud mediante una combinación de seis dimensiones: “Funcionamiento físico” (FF), “Limitaciones del rol” (LR), “Funcionamiento social” (FS), “Dolor” (DO), “Salud mental” (SM) y “Vitalidad” (VI); cada una de las cuales ha adoptado un número variable de niveles. Los análisis de componentes principales realizados por Brazier et al. (1998, 2002) les permitieron reducir el número de dimensiones del SF-36 de 8 a 6, mediante la exclusión del atributo “Salud general” y la fusión de las dos dimensiones de Limitaciones en el rol (físico y emocional). Los niveles correspondientes a cada una de las seis dimensiones del SF-6D provienen de una selección de 11 ítems de los 36 que contiene el SF-36.⁴

⁴ Existe una versión diferente del SF-6D (Brazier y Roberts, 2004), cuya singularidad radica en que está derivada a partir del SF-12 y no del SF-36, y que contempla sólo 3 niveles en “Funcionamiento físico” y únicamente cinco niveles en “Dolor”, en lugar de los seis del SF-6D(SF-36).

Figura 1. El sistema de clasificación de estados de salud SF-6D

Funcionamiento físico (FF)

1. Su salud no le limita para realizar esfuerzos intensos (p.ej. correr, levantar objetos pesados, participar en deportes agotadores).
2. Su salud le limita un poco para realizar esfuerzos intensos (p.ej. correr, levantar objetos pesados, participar en deportes agotadores).
3. Su salud le limita un poco para realizar esfuerzos moderados (p.ej. mover una mesa, pasar la aspiradora o caminar más de 1 hora).
4. Su salud le limita mucho para realizar esfuerzos moderados (p.ej. mover una mesa, pasar la aspiradora o caminar más de 1 hora).
5. Su salud le limita un poco para bañarse o vestirse por sí mismo.
6. Su salud le limita mucho para bañarse o vestirse por sí mismo.

Limitaciones de rol (LR)

1. No tiene problemas con su trabajo u otras actividades cotidianas a causa de su salud física o de sus problemas emocionales.
2. Está limitado en su trabajo o en sus actividades cotidianas a causa de su salud física.
3. Hace menos de lo que quisiera hacer a causa de sus problemas emocionales.
4. Está limitado 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.

Funcionamiento social (FS)

1. Su salud no le dificulta sus actividades sociales (como visitar a amigos o familiares) en ningún momento.
2. Su salud le dificulta sus actividades sociales (como visitar a amigos o familiares) sólo alguna vez.
3. Su salud le dificulta sus actividades sociales (como visitar a amigos o familiares) algunas veces.
4. Su salud le dificulta sus actividades sociales (como visitar a amigos o familiares) casi siempre.
5. Su salud le dificulta sus actividades sociales (como visitar a amigos o familiares) siempre.

Dolor (DO)

1. No tiene dolor.
2. Tiene dolor pero no interfiere con su trabajo habitual (fuera de casa o en las tareas del hogar).
3. Tiene dolor que interfiere un poco con su trabajo habitual (fuera de casa o en las tareas del hogar).
4. Tiene dolor que interfiere moderadamente con su trabajo habitual (fuera de casa o en las tareas del hogar).
5. Tiene dolor que interfiere bastante con su trabajo habitual (fuera de casa o en las tareas del hogar).
6. Tiene dolor que interfiere mucho con su trabajo habitual (fuera de casa o en las tareas del hogar).

Salud mental (SM)

1. Nunca se siente muy nervioso o desanimado y deprimido.
2. Se siente muy nervioso o desanimado y deprimido sólo alguna vez.
3. Se siente muy nervioso o desanimado y deprimido algunas veces.
4. Se siente muy nervioso o desanimado y deprimido casi siempre.
5. Se siente muy nervioso o desanimado y deprimido siempre.

Vitalidad (VI)

1. Tiene mucha energía siempre.
2. Tiene mucha energía casi siempre.
3. Tiene mucha energía algunas veces.
4. Tiene mucha energía sólo alguna vez.
5. Nunca tiene mucha energía.

Con la combinación de atributos y niveles del SF-6D que se muestran en la Figura 1 es posible definir un total de 18000 estados de salud diferentes.

2.1.3. Escalas multiatributo: los algoritmos o funciones de utilidad

El principal atractivo de los tres instrumentos que se han descrito en el apartado anterior radica, como se ha dicho, en que para todos ellos se dispone de, al menos, una tarifa o panel de utilidades asociadas a los estados de salud que cada sistema es capaz de caracterizar; de modo que es posible determinar la medida cardinal de la utilidad asociada a cada estado de salud descrito mediante cualquiera de estos tres sistemas, mediante la aplicación de un algoritmo de cálculo, generalmente sencillo.

La tarea de generar una tarifa o panel de utilidades completo para un sistema de clasificación multiatributo conlleva la estimación de un modelo, para lo cual, en general, se puede partir de dos enfoques alternativos (Dolan, 2002; Stevens et al., 2007): uno es el denominado enfoque compuesto o de inferencia estadística, mientras que el otro es el enfoque descompuesto o algebraico, basado en la teoría de la utilidad multiatributo.

El enfoque compuesto requiere que cada encuestado valore un subconjunto de estados de salud multiatributo. El algoritmo se estima mediante la aplicación de técnicas de regresión lineal, por lo general con modelos de efectos aleatorios. Esta aproximación, más preocupada por la validez empírica que por la teórica⁵, es la adoptada en la inmensa mayoría de las estimaciones realizadas con el EQ-5D (Dolan, 1997; Badía et al., 2001; Tsuchiya et al., 2002; Lamers et al., 2006) y con el SF-6D (Brazier et al., 1998; 2002; 2004; Lam et al., 2008; Abellán et al., 2008, 2010), y sólo en una ocasión, hasta donde sabemos, con el HUI (McCabe et al., 2005a).

El enfoque algebraico requiere que los encuestados valoren cada nivel o ítem de un atributo determinado manteniendo constantes los niveles del resto de atributos (por lo general en el mejor nivel). Las funciones de utilidad uni-atributo resultantes se utilizan para generar las utilidades de estados de salud compuestos (o multiatributo) especificando una función de utilidad multiatributo. Este enfoque es el que se ha utilizado generalmente para estimar los algoritmos del HUI (Torrance et al., 1982, 1996; Feeny et al., 2002; Mc Cabe et al. 2005b) y es el que se asume en la presente investigación. Su principal ventaja radica en que minimiza el número de estados de

⁵ En el sentido de que la función que se estima estadísticamente es *ad hoc*, ya que no se hace explícita la función de utilidad multiatributo que subyace a la misma.

salud que es necesario valorar de manera directa, para luego poder realizar predicciones *out-of-sample*.

Stevens et al. (2007) compararon los resultados obtenidos en McCabe et al. (2005a,b) al aplicar los dos enfoques a un mismo instrumento, el HUI2, concluyendo que la validez predictiva (esto es, la capacidad que tienen las utilidades predichas por los modelos de coincidir con las utilidades valoradas directamente por la muestra) conseguida con el enfoque de inferencia estadística era superior a la registrada mediante el enfoque de la función de utilidad multiatributo.

Puesto que en otra parte (Abellán et al., 2008) se ha descrito con suficiente detalle la metodología propia del enfoque “compuesto” o de inferencia estadística, en el apartado siguiente se expone el procedimiento de modelización que sigue el enfoque “descompuesto” o algebraico, cuyas raíces están en la teoría de la utilidad multiatributo. Así mismo, se da cuenta de la aplicación de dicho enfoque al sistema descriptivo HUI que, hasta donde sabemos, es el único instrumento multiatributo cuya función de utilidad ha sido modelizada siguiendo esta metodología.

2.2. La obtención de funciones de utilidad multiatributo a partir del enfoque descompuesto o algebraico.

2.2.1. El modelo multiatributo

La teoría de la utilidad multiatributo (Keeney and Raiffa, 1976) proporciona un método eficiente para determinar las utilidades de los estados de salud descritos mediante sistemas de clasificación como el SF-6D, el EQ-5D o el HUI. El enfoque consiste básicamente en medir las funciones de utilidad de cada uno de los atributos o dimensiones de que consta el instrumento, considerados dichos atributos de forma separada (funciones “uni-atributo”) y determinar, seguidamente, una ecuación que exprese la utilidad total como una función de estas utilidades uni-atributo.

Existen formas funcionales alternativas (aditiva-lineal, multiplicativa, multilineal) y la teoría especifica las condiciones bajo las cuales cada una de estas formas funcionales resulta apropiada (independencia aditiva, independencia mutua de la utilidad, e independencia de primer orden, respectivamente).

Dado que la elección de una forma funcional condiciona también el número de parámetros y, por ende, el número de medidas directas de utilidad multiatributo (por

oposición a uni-atributo) necesarias, la forma multilineal no parece, a priori, una alternativa plausible, por requerir un número de parámetros muy elevado y, consiguientemente, la medida directa de la utilidad multiatributo de una cantidad inasumible de estados de salud.

Por otra parte, puesto que el modelo aditivo-lineal es un caso particular del multiplicativo, una buena opción consiste en obtener las medidas multiatributo necesarias para ajustar este último. La forma general de una función de utilidad multiatributo multiplicativa con n atributos es la siguiente (Torrance et al., 1982; 1996):

$$u = \frac{1}{k} \left[\prod_{j=1}^n 1 + k k_j u_j - 1 \right] \quad (1)$$

Donde

$$1 + k = \prod_{j=1}^n 1 + k k_j \quad (2)$$

En la expresión (1), u es la utilidad cardinal del estado de salud en la escala “peor-mejor” estado de salud, es decir, la escala en la que el estado *all worst* tiene valor cero y el *all best* tiene valor uno. Por otra parte, u_j es la utilidad uni-atributo de la dimensión j , en la escala en la que el peor nivel del atributo j tiene una utilidad cero y el mejor nivel una utilidad igual a la unidad. Finalmente, k y k_j son constantes o parámetros de escalamiento de la función de preferencias que se estiman a partir de los datos. Las ecuaciones anteriores pueden (y de hecho, así suele hacerse) ser expresadas en términos de desutilidad (\bar{u}):

$$\bar{u} = \frac{1}{c} \left[\prod_{j=1}^n 1 + c c_j \bar{u}_j - 1 \right] \quad (3)$$

$$1 + c = \prod_{j=1}^n 1 + c c_j \quad (4)$$

Siendo $\bar{u} = 1 - u$, y $\bar{u}_j = 1 - u_j$.

Para determinar las seis constantes c_j de la ecuación 6 (asumiendo la formulación en términos de desutilidad), se necesita conocer las utilidades (o desutilidades) de seis

estados de salud multiatributo medidas en una escala cuyos extremos sean “salud normal” y “el peor estado de salud”. La selección de los estados debe permitir la obtención de todas las constantes y además estos estados deben ser realistas. La alternativa más sencilla es seleccionar los estados “esquina” (*corner states*), definidos como aquéllos en los que solo un atributo se sitúa en su peor nivel mientras que los restantes se mantienen en su mejor nivel (nivel 1).

Esto simplifica enormemente el procedimiento, pues el valor de c_j en la ecuación (3) se iguala simplemente a la desutilidad del estado “esquina” correspondiente a la dimensión j . Sin embargo, se puede plantear el problema de que algunos de estos estados esquina pueden resultar poco plausibles a ojos del encuestado, por ser considerados escasamente realistas. Para el caso concreto del SF-6D, por ejemplo, es difícil imaginar a alguien en el peor nivel posible de “funcionamiento físico” (muy limitado para bañarse o vestirse por sí mismo) y, simultáneamente, en el mejor nivel de la dimensión “limitaciones del rol” (sin problemas en el trabajo u otras actividades cotidianas).

Para evitar este problema de falta de verosimilitud puede recurrirse a estados “cuasi-esquina” o “retirados” (*backed-off corners*), entendiéndose por tales aquéllos en los que todas las dimensiones están en el nivel 1 (el mejor) excepto una que se encuentra en un nivel próximo al máximo (peor) pero sin llegar a alcanzarlo; así como a estados “doble-esquina” (*double corner state*) en los que dos dimensiones se sitúan en su nivel máximo y el resto en el nivel 1.

Las desutilidades de las dimensiones para las que se ha definido un estado “cuasi-esquina” se infieren extrapolando desde la posición “retirada” hasta la “esquina”, a partir de la función de preferencia uni-atributo del individuo. Las dimensiones para las que se hayan definido estados “doble-esquina” se tratan de manera diferente, toda vez que el recurso a un estado “doble-esquina” refleja la ausencia de independencia estructural entre ambos atributos (Torrance et al., 1996).

Una vez que se hayan determinado las constantes c_j de la ecuación (3), el valor de c se obtiene mediante la resolución iterativa de la ecuación (4). La suma de las seis constantes c_j indica hacia dónde se ha de buscar la solución iterativa: si la suma es

mayor que 1, entonces $-1 < c < 0$; si la suma es igual a 1, entonces $c = 0$ (modelo aditivo-lineal); y si la suma es menor que 1, entonces $c > 0$.

Obtenidos todos los parámetros del modelo (los valores de c y de cada c_j), la expresión (3) permite calcular la desutilidad de cada estado de salud multiatributo, en la escala en la que el valor 0 corresponde al peor estado de salud (*all worst*) y el valor 1 al estado de perfecta salud. Para derivar las utilidades medidas en una escala en la que el 0 se identifique con la muerte, será preciso convertir las desutilidades a utilidades ($u = 1 - \bar{u}$) y reescalarlas atendiendo a la utilidad asignada a la muerte en la escala multiatributo donde el peor estado de salud ocupa el extremo inferior del intervalo.

2.2.2. El enfoque algebraico en la práctica: funciones de utilidad del Health Utilities Index

La aproximación algebraica a la obtención de funciones de utilidad multiatributo, con las que poder generar tarifas o paneles de utilidades, se ha utilizado hasta la fecha únicamente con uno de los instrumentos que han sido mencionados en las páginas anteriores, el *Health Utilities Index* (HUI).

La primera aplicación de la teoría de la utilidad multiatributo para medir preferencias sociales por estados de salud se debe a Torrance et al. (1982). En este estudio, el enfoque algebraico descrito en el apartado precedente se aplicó a un sistema de clasificación de estados de salud con cuatro atributos, que posteriormente se conocería como la versión 1 del HUI (*Health Utilities Index Mark 1*). Para la obtención de las funciones de utilidad de cada dimensión ('uniatributo') recurrieron al método de la escala visual analógica, transformando los valores (v) así obtenidos en utilidades (u) mediante una transformación potencial basada en estudios previos⁶:

$$u = (v)^{1.6}$$

Para las medidas multiatributo (los cuatro estados esquina "puros" y el peor estado de salud, además de otros dos estados que servirían para testar el ajuste del modelo), se recurrió al método de la equivalencia temporal o *time trade-off* (TTO). De sus resultados se desprende que el modelo multiplicativo era más apropiado que el

⁶ Con el término 'utilidades' se hace referencia a índices cardinales de preferencias que incorporan la actitud hacia el riesgo de los encuestados, en oposición a los 'valores', que son obtenidos en un contexto de certeza.

aditivo-lineal para todos los sujetos y para el agregado, lo que, advierten, no tiene por qué ser generalizable a cualquier instrumento de clasificación multiatributo.

Torrance et al. (1996) obtuvieron la función de utilidad de la segunda versión del instrumento, el *HUI Mark 2* (o HUI:2). Una primera diferencia respecto del estudio de Torrance et al. (1982) radica en los métodos de valoración. De un lado, las utilidades multiatributo se midieron a través de la “lotería estándar” (*standard gamble*), en lugar de con el TTO. Por otra parte, aunque las utilidades uniatributo se derivaron siguiendo un procedimiento similar al del trabajo original, esto es, midiendo los “valores” dentro de cada dimensión mediante una escala visual y convirtiéndolos posteriormente a utilidades mediante una función de transformación potencial, la diferencia estriba en que el parámetro de actitud hacia el riesgo se obtuvo ajustando una curva a las valoraciones y utilidades que, para unos mismos estados de salud multiatributo (cuatro en total) se habían obtenido en la encuesta mediante la escala visual y la lotería estándar, respectivamente. La función se ajustó en términos de desutilidad, siendo su expresión en utilidades (u)/valores (v):

$$u = 1 - (1-v)^{2.29}$$

Puesto que, para cada individuo, se dispuso de las valoraciones uniatributo y de las utilidades multiatributo, los autores pudieron estimar funciones de utilidad individuales (con cuyas predicciones se calcularon las utilidades medias predichas para cada estado de salud), así como la función de utilidad basada en las valoraciones medias, que resultó ser el modelo recomendado para su uso debido a la mayor simplicidad inherente al enfoque. Al igual que en el caso del HUI:1, los resultados rechazaron el modelo lineal.

Con una metodología muy similar, Feeny et al. (2002) derivaron la función de utilidad para el *HUI Mark 3*. La principal diferencia respecto al estudio anterior estriba en el tratamiento que se dio a la valoración de estados considerados peores que la muerte (Torrance et al., 1996 recurrieron sólo a la escala visual, mientras que Feeny et al. 2002 valoraron también estos estados mediante un método similar al *standard gamble* el *chance board*). A esta diferencia atribuyen, en parte, Feeny et al. el hecho de que el valor mínimo que resulta de su algoritmo (-0.36) sea notablemente inferior al de la



función estimada para el HUI:2 (-0.03). Un resultado que sí coincide con estimaciones de las versiones anteriores del HUI es el firme rechazo de la forma funcional lineal.

Mencionaremos para terminar, brevemente, el trabajo de McCabe et al. (2005b), quienes replicaron la modelización de Torrance et al. (1996) para el HUI:2 con una muestra de población del Reino Unido y ensayaron una función de transformación de las valoraciones de la escala visual a utilidades *standard gamble* cúbica, como alternativa a la función potencial de Torrance y otros, concluyendo que la función multiplicativa multiatributo cúbica era mejor en cuanto a su capacidad predictiva que la potencial.

3. Motivación y objetivos del estudio.

3.1. Motivación

El presente estudio comparte con el que se desarrolló al amparo del primer Convenio de colaboración en materia de Economía de la Salud suscrito entre la Universidad y la C.A.R.M. (Abellán et al., 2008), la motivación fundamental de disponer de un algoritmo de puntuación para el instrumento conocido como SF-6D, que permitiese generar pesos de CVRS y que fuese, en consecuencia, susceptible de ser aplicado en el cálculo de Años de Vida Ajustados por la Calidad (AVAC) necesario para la realización de estudios de evaluación económica de nuevas tecnologías sanitarias.

Dado que la primera aproximación a esta tarea optó por uno de los dos enfoques metodológicos al uso, a saber, el basado en inferencia estadística, el equipo investigador consideró conveniente completar dicha aproximación con la resultante del otro enfoque posible, esto es, el fundamentado en la teoría de la utilidad multiatributo o enfoque “descompuesto”. La finalidad de ello es doble: de un lado, se trataría de ensayar con el SF-6D una aproximación metodológica inédita hasta la fecha; de otro lado, los resultados podrían ser comparados con los obtenidos en el primer estudio de manera que se testase la robustez de las estimaciones y se identificasen, en su caso, las fuentes de discrepancia entre ambos enfoques, con la intención de seleccionar la mejor estrategia para la obtención de una “tarifa” o panel de utilidades del instrumento SF-6D.

Hasta donde llega nuestro conocimiento, únicamente existe un precedente en el que se haya abordado la obtención del algoritmo para un instrumento de CVRS multiatributo basado en preferencias a través de los dos enfoques alternativos antes citados. Nos referimos al trabajo de Stevens et al. (2007), quienes confrontan los resultados derivados de cada una de las aproximaciones (McCabe et al., 2005a y McCabe et al., 2005b) aplicadas al HUI:2, concluyendo que la validez predictiva del enfoque de inferencia estadística era claramente superior a la derivada de aplicar el enfoque de la función de utilidad multiatributo.

3.2. Objetivos

El objetivo principal del proyecto es la estimación empírica de un algoritmo para el instrumento descriptivo SF-6D, basada en la teoría de la utilidad multiatributo. Dicho algoritmo permitirá asignar a cada condición o problema de salud construido a partir de las respuestas al cuestionario SF-36 (a partir del cual se “extrae” la codificación según el sistema SF-6D), un valor representativo de las preferencias de la población general murciana. En consecuencia, el algoritmo estimado predecirá las utilidades de los estados SF-6D, de modo que tales utilidades puedan emplearse en diferentes contextos como, por ejemplo, la evaluación económica de intervenciones sanitarias, el análisis coste-beneficio generalizado del sistema murciano de salud, o simplemente el cómputo del stock de salud de la población murciana. Objetivos secundarios del proyecto son varios:

1. El perfeccionamiento metodológico del procedimiento de estimación del algoritmo. Concretamente, se recurrirá a una técnica de obtención de preferencias basada en elecciones, en lugar del procedimiento en dos etapas habitual en este enfoque, según el cual las funciones de utilidad de cada atributo por separado (funciones “uniatributo”), así como las valoraciones de los estados multiatributo se obtienen mediante el método de la escala visual y, posteriormente se “transforman” en utilidades mediante una función potencial. Concretamente, en nuestro estudio utilizaremos una técnica de ‘doble lotería’ (Pinto y Abellán, 2005): el método *probability lottery equivalence* (Bleichrodt et al., 2007).
2. La comparación de los resultados obtenidos al estimar el algoritmo del SF-6D mediante el enfoque algebraico, con los que se obtuvieron mediante técnicas de inferencia estadística, con el fin de concluir cuál de las dos aproximaciones metodológicas ofrece un mejor balance en términos, no sólo de capacidad predictiva o “bondad de ajuste”, sino también de parsimonia y factibilidad.
3. Recopilar información cualitativa sobre la salud de la población general murciana, a partir de las puntuaciones otorgadas en el SF-36 a cada una de sus ocho dimensiones, así como sobre la base de sus respuestas al autoclasificador EQ-5D. Dicho acopio de información resultará útil para profundizar en el conocimiento de la calidad de vida de los habitantes de la región.

4. Material y métodos.

4.1. Modelización y procedimientos de medición de las utilidades

4.1.1. La función de utilidad multiatributo del SF-6D

La teoría de la utilidad multiatributo proporciona un método eficiente para determinar las utilidades de los 18000 estados de salud que el SF-6D permite describir a partir de la combinación de sus seis dimensiones y los niveles (entre 4 y 6) definidos para cada dimensión (ver Tabla 1). Como se ha señalado en el capítulo 2, el enfoque consiste en medir las seis funciones de utilidad de cada atributo considerado de forma separada (funciones “uni-atributo”) y determinar una ecuación que exprese la utilidad total como una función de estas utilidades uni-atributo.

Por las razones expuestas en el apartado 2.2, nos decidimos por el modelo multiplicativo, cuya función de utilidad con seis atributos es la siguiente:

$$u = \frac{1}{k} \left[\prod_{j=1}^6 1 + k k_j u_j - 1 \right] \quad (5)$$

Donde

$$1 + k = \prod_{j=1}^6 1 + k k_j \quad (6)$$

En la expresión (5), u es la utilidad cardinal del estado de salud en la escala “peor-mejor” estado de salud, es decir, la escala en la que el estado *all worst* (645655) tiene valor cero y el *all best* (111111) tiene valor uno. Por otra parte, u_j es la utilidad uni-atributo de la dimensión j , en la escala en la que el peor nivel del atributo j tiene una utilidad cero y el mejor nivel una utilidad igual a la unidad. Finalmente, k y k_j son constantes o parámetros de escalamiento de la función de preferencias que se estiman a partir de los datos. Las ecuaciones anteriores pueden (y de hecho, así suele hacerse) ser expresadas en términos de desutilidad (\bar{u}):

$$\bar{u} = \frac{1}{c} \left[\prod_{j=1}^6 1 + c c_j \bar{u}_j - 1 \right] \quad (7)$$

$$1 + c = \prod_{j=1}^6 1 + cc_j \quad (8)$$

Siendo $\bar{u} = 1-u$, y $\bar{u}_j = 1-u_j$.

Por ejemplo, la función de desutilidad uni-atributo para la dimensión “Dolor”, u_4 , se determina en una escala de desutilidad en la que el mejor nivel (nivel 1) tiene una desutilidad de 0.0 y el peor nivel (nivel 6) tiene una desutilidad de 1.0. Durante el proceso de obtención de la función de utilidad uniatributo, se pide al entrevistado que imagine que los restantes cinco atributos no cambian. De igual modo se obtienen las funciones de utilidad uni-atributo de las otras cinco dimensiones del SF-6D.

Para determinar las seis constantes c_j de la ecuación 7 (asumiendo la formulación en términos de desutilidad), se necesita conocer las desutilidades de seis estados de salud multiatributo medidas en una escala cuyos extremos sean “salud normal” y “el peor estado de salud”. La selección de los estados debe permitir la obtención de todas las constantes y además estos estados deben ser realistas. La alternativa más sencilla e, como se ha dicho más arriba, seleccionar los seis estados “esquina” (*corner states*), definidos como aquéllos en los que solo un atributo se sitúa en su peor nivel mientras que los cinco restantes se mantienen en su mejor nivel (nivel 1).

Sin embargo, en la sección 2 se comentó el problema de falta de verosimilitud que podrían plantear algunos estados esquina “puros”. Así, de manera análoga a lo que hicieron Torrance et al. (1996) en el caso del *HUI:2*, en nuestro estudio incluiremos tres estados “esquina” en sentido estricto (*pure corner states*) para las dimensiones “Funcionamiento social” y “Vitalidad”, tres estados “cuasi-esquina” o “retirados” (*backed-off corners*) para los atributos “Funcionamiento físico”, “Dolor” y “Salud mental”, y un estado “doble-esquina” (*double corner state*) para el atributo “Limitaciones del rol”, que se dobla con “Funcionamiento físico”. En consecuencia, necesitaremos disponer de las utilidades multiatributo de los estados: (641111), (411111), (115111), (111411), (111131) y (111115), además del estado *all worst*, que también recibe la denominación de *pits state* (645655), tal y como se muestra en la Tabla 1. La utilidad de este último estado se necesita para poder cambiar la escala de las desutilidades cuando sea necesario, de una escala en la que el valor cero lo ocupe

la muerte a otra escala en la que dicho valor esté asignado al *pits state* y viceversa, con el fin de poder determinar los parámetros de la función multiatributo expresada en la ecuación (7) y disponer, al final del proceso, de un algoritmo que nos permita obtener directamente las utilidades de los estados de salud en una escala (0,1) en la que el extremo inferior sea la muerte y el extremo superior la salud normal.

Tabla 1. Estados de salud multiatributo valorados directamente

<i>Estados esquina 'puros'</i>	<i>Estados 'cuasi-esquina' (backed-off)</i>	<i>Estados 'doble-esquina'</i>	<i>El peor estado de salud</i>
115111 111115	411111 111411 111131	641111	645655

Como se ha señalado antes, las desutilidades de los estados esquina 'puros' proporcionarán de manera directa los valores de c_3 y c_6 (correspondientes a las dimensiones "Funcionamiento social" y "Vitalidad"). Por su parte, las desutilidades de los atributos "Dolor" y "Salud mental" (y, consiguientemente, los valores de c_4 y c_5) se inferirán extrapolando desde la posición "retirada" hasta la "esquina", a partir de la función de preferencia uni-atributo del individuo. Las dimensiones "Funcionamiento físico" y "Limitaciones del rol" se tratarán de manera diferente, toda vez que están valoradas mediante un estado "doble-esquina",

Una vez que se hayan determinado las constantes c_j de la ecuación (7), el valor de c se obtendrá mediante la resolución iterativa de la ecuación (8). La suma de las seis constantes c_j indica hacia dónde se ha de buscar la solución iterativa: si la suma es mayor que 1, entonces $-1 < c < 0$ lo que significará que los atributos son complementarios. Si la suma de las c_j es menor que 1, entonces $c > 0$ lo que indicará una relación de sustituibilidad entre dimensiones. Finalmente, si la suma de los parámetros de escalamiento es igual a 1, entonces $c = 0$, lo que implicará que no existen interacciones en las preferencias entre atributos, obteniéndose una función de utilidad aditiva lineal.

4.1.2. Obtención de las utilidades: método de elicitación

Los estudios previos utilizaron el método de la "equivalencia temporal" o TTO (Torrance et al., 1982), o bien la "lotería estándar" (Torrance et al., 1996; Feeny et al., 2002) para medir utilidades multiatributo. Por otro lado, tanto las utilidades

uniatributo como las de los estados multiatributo valorados directamente (los utilizados en la modelización –*corner*, *backed-off*, *double corner*– y los posteriormente empleados para valorar el ajuste del modelo) fueron obtenidas de modo indirecto, a saber, mediante la transformación con una función potencial de las valoraciones asignadas en una escala visual analógica. En nuestro estudio, y esta es una de las principales diferencias respecto de los citados, medimos directamente utilidades y, además, empleamos para ello el mismo método tanto en el caso de las funciones uniatributo como en el de los estados multiatributo de la tabla 1.

El método elegido es un tipo concreto de “doble lotería” o “lotería equivalente” (McCord y De Neufville, 1986). Concretamente, dentro de las técnicas de “doble lotería”, hemos optado por el procedimiento denominado *probability lottery equivalence* (PLE), según la terminología utilizada en Bleichrodt et al. (2007). La diferencia entre esta técnica y la denominada *value lottery equivalence* (VLE) radica en el parámetro con el que se logra la indiferencia: en el caso de la PLE dicho parámetro es la probabilidad, mientras que en la VLE es otro distinto como, por ejemplo, la duración asociada a un estado de salud.

En esencia, el método de elicitación que aplicamos en la valoración de los estados multiatributo pasa por encontrar la distribución de probabilidades que hace que el sujeto sea indiferente entre dos “loterías”: una en la que los posibles sucesos son la salud normal y la muerte; y otra en la que los *outcomes* son la salud normal y el estado crónico *H* cuya utilidad se pretende obtener. El diseño concreto que hemos aplicado consiste en mantener constantes las probabilidades de la lotería en la que el estado *H* es uno de los resultados posibles y buscar la indiferencia con la probabilidad asociada a la salud normal en la lotería donde está presente la muerte como probable *outcome*. Se trata, en consecuencia, de hallar el valor de p que consigue la indiferencia entre las dos loterías⁷:

$$(SN, p, Muerte) \sim (SN, 0.5, Estado H) \quad (9)$$

En la mayor parte de los casos se obtiene que $p > 0.5$ lo que indica que el sujeto admite un cierto riesgo de morir $(1-p) < 0.5$ a cambio de evitar un riesgo del 50% de sufrir el

⁷ Ver Apéndice 1 para una descripción más detallada de los escenarios de valoración y del tipo de preguntas que se plantean a los entrevistados.

estado H de manera crónica. Sin embargo, algunos sujetos pueden considerar que uno o varios de los estados de salud, por su especial gravedad, son “peores que la muerte”, en cuyo caso se obtendrán valores de p inferiores al 50%. Dado que se ha utilizado un cuestionario asistido por ordenador, y puesto que la aplicación informática incluye un procedimiento secuencial de convergencia al valor de indiferencia (ver Apéndice), se ha incluido una pregunta inicial que nos permite saber si el encuestado considera el estado Q mejor o peor que la muerte (aunque el *framing* del método es idéntico en ambos casos, lo que varía en uno u otro supuesto es el sentido ascendente o descendente en el que debe variarse el estímulo p a la vista de las respuestas del sujeto).

Para la obtención de las funciones de utilidad uni-atributo, el método es igual al descrito, salvo por el hecho de que, en este caso, lo que se valora no es un estado de salud H sino cada nivel i intermedio (es decir, distinto del peor y del mejor nivel) dentro de una dimensión j . Consiguientemente, en el *framing* de la “doble lotería” la salud normal es reemplazada por el mejor nivel o nivel 1 en dicha dimensión j (esto es, la ausencia total de problemas en el atributo objeto de medición, mientras que, en la lotería que se utiliza como estímulo, el resultado adverso no es en este caso la muerte, sino el nivel máximo (peor) en la escala del atributo en cuestión. Así, dentro de cada uno de los seis atributos o dimensiones j las utilidades de los niveles intermedios i se derivan a partir de los valores de p que hacen indiferentes a los sujetos entre las siguientes dos loterías:

$$(\text{Nivel } 1, p, \text{ Nivel } M) \sim (\text{Nivel } 1, 0.5, \text{ Nivel } i) \quad (10)$$

Siendo M el peor nivel o nivel máximo alcanzable (4, 5 o 6, según de qué atributo se trate) en la dimensión cuya función de utilidad uniatributo se evalúa.

En el caso de las valoraciones dentro de un atributo, no ha lugar a la consideración de una determinada situación como “peor que la muerte”, pues este resultado no forma parte del espacio de posibles *outcomes*. Lo que sí sería posible, al menos en teoría, es que un sujeto proporcionase un valor de indiferencia para p menor que 0.5. Sin embargo, excluimos a priori la existencia de preferencias monótonas crecientes en la gravedad (p.ej. preferir más dolor a menos dolor), por lo que en caso de que un sujeto hubiera tenido la oportunidad de aceptar una probabilidad inferior a 0.5 sería

simplemente consecuencia de un error. Por esta razón, la aplicación informática se modificó respecto de la utilizada para las valoraciones multiatributo, restringiendo los posibles valores de p a aquellos comprendidos entre 0.5 y 1.⁸

4.1.3. Cálculo de las utilidades

Si identificamos por p^* al valor que consigue la indiferencia en la expresión (9), siendo $U(SN)=1$ y $U(Muerte)=0$, y asumiendo que se verifican los supuestos de la utilidad esperada, podemos deducir de la citada expresión que $p^* + 0 = 0.5 + 0.5 U(Q)$, de donde se obtiene el siguiente valor para la utilidad del estado de salud H :

$$U(H) = 2p^* - 1 \quad (11)$$

El cálculo de la utilidad del estado de salud valorado con el método PLE es idéntica con independencia de que el estado se considere mejor o peor que la muerte. Una ventaja adicional de la forma elegida para administrar el método de elicitación (al margen de la ya argumentada superioridad del procedimiento frente a la lotería estándar), radica en que las utilidades que se obtienen para estados peores que la muerte están acotadas entre 0 y -1 , no siendo necesario aplicar método alguno de reescalamiento, con lo que se evita la arbitrariedad que este tipo de procedimientos conlleva en mayor o menor grado.

En el caso de las utilidades uniatributo, el cálculo de la utilidad es idéntico al mostrado en la expresión (11), solo que ahora los extremos del intervalo son diferentes, siendo 1 la utilidad asociada al mejor nivel de la correspondiente dimensión (nivel 1) y 0 la utilidad que se asigna al peor nivel definido para dicho atributo (4, 5 o 6, según el caso).

Todas las utilidades se convertirán a desutilidades ($\bar{u} = 1-u$, y $\bar{u}_j = 1-u_j$). Adicionalmente, para poder determinar las constantes c_j de la ecuación (7), será

⁸ Únicamente para la dimensión 1, "Funcionamiento físico", la preferencia por un nivel inferior al nivel 6 ("Su salud le limita mucho para bañarse o vestirse por sí mismo") puede no interpretarse como una inconsistencia lógica, en la medida en que un nivel como el 4 ("Su salud le limita mucho para realizar esfuerzos moderados") puede ser considerado por un individuo como un escenario peor o menos deseable que el descrito en el nivel 6. No obstante, los resultados de un pre-test nos indicaron que una respuesta de este tipo era muy poco probable, como posteriormente se confirmó en la encuesta definitiva: antes de enfrentarse a las preguntas con el método PLE los entrevistados debían situar los diferentes niveles en una escala visual cuyos extremos eran el mejor y el peor nivel de la dimensión evaluada, no registrándose ningún caso en el que los niveles 4 o 5 de la dimensión 1 se situasen, no ya en el extremo inferior, sino ni siquiera por debajo de la marca correspondiente al 10 en la escala 0-100.

preciso convertir las desutilidades de los seis estados multiatributo (medidas en una escala en la que el 0 se identifica con la muerte) a la escala cuyos extremos son “salud normal” y “el peor estado de salud”. Para ello, las desutilidades se dividirán entre la desutilidad del *pits state*.

4.2. Recogida de datos

4.2.1. Las muestras

Las utilidades de los estados multiatributo recogidos en la Tabla 1 y las funciones de utilidad uniatributo para cada dimensión se obtuvieron a partir de muestras de población diferentes. La primera muestra (muestra 1) coincide con la utilizada para la obtención del algoritmo del SF-6D según el método compuesto.⁹ Así, los estados multiatributo que en el presente estudio sirven para la obtención de la tarifa según el enfoque descompuesto, formaron parte de la selección de 78 estados de salud que fueron objeto de medición directa en la aproximación basada en inferencia estadística.

Tal y como consta en el informe el estudio precedente, entonces se aplicó un diseño “entre-muestras”, esto es, los estados se repartieron en subconjuntos de cinco que se distribuyeron para su valoración por subgrupos de la muestra total. De los 17 subgrupos en los que se dividió la muestra, 14 de ellos hubieron de valorar, junto con otros cuatro estados SF-6D, alguno de los 7 estados cuyas utilidades se emplearán para la obtención de la tarifa SF-6D a través del enfoque algebraico. Teniendo en cuenta que cada una de las submuestras tenía un tamaño de 60 individuos y que cada uno de los siete estados de la Tabla 1 se asignó a dos submuestras distintas, el tamaño total de la muestra 1, esto es, del conjunto de participantes en la primera encuesta cuyas respuestas han sido tenidas en cuenta en el presente estudio es de 840 sujetos, siendo 120 el número de valoraciones por cada estado de salud. Puesto que se garantizó la representatividad poblacional por sexo y por cuotas de edad dentro de cada uno de los grupos de 60 individuos, el conjunto de valoraciones de cada estado de salud multiatributo utilizado en este estudio procede de una muestra representativa de la población adulta de la Región de Murcia por sexo y grupos de edad.

⁹ Ver Abellán et al. (2008).

Para la obtención de las funciones de utilidad uniatributo se contó con una muestra de 390 individuos extraídos de la población adulta de la Región de Murcia (muestra 2), distribuidos en tres sub-grupos cada uno de ellos representativos de la población según sexo y grupos de edad. También en este caso se aplicó un diseño “entre muestras”, con el fin de evitar los posibles errores o sesgos en las respuestas que pudieran derivarse del cansancio que podría ocasionar en el sujeto la realización de un número excesivo de tareas reiterativas. De este modo, cada entrevistado respondió a las preguntas de valoración de niveles correspondientes a 2 de las 6 dimensiones del instrumento SF-6D. Concretamente, de una de las submuestras se obtuvieron las funciones de utilidad de los atributos 1 (“Funcionamiento físico”) y 5 (“Salud mental”), de otra de ellas las correspondientes a los atributos 2 (“Limitaciones de rol”) y 4 (“Dolor”), y de la tercera las utilidades uniatributo para las dimensiones 3 (“Funcionamiento social”) y 6 (“Vitalidad”). En consecuencia, el número de valoraciones por nivel/dimensión en esta segunda muestra fue de 130.¹⁰

4.2.2. Los cuestionarios

Ambos cuestionarios se estructuraron en cuatro partes, la primera de las cuales se dedicó a explicar a los sujetos las características del sistema de clasificación de estados de salud SF-6D; en el caso de la muestra 2 únicamente en lo relativo a las dos dimensiones incluidas en el modelo de cuestionario que hubiese correspondido al encuestado. Describimos a continuación, brevemente, el contenido de las tres secciones restantes de los cuestionarios, remitiendo al lector al Apéndice 1 para una exposición más detallada.

En la parte 2, los sujetos que integraron la muestra 1 tuvieron que situar cinco estados SF-6D sobre un “termómetro” rotulado de 0 (*el peor estado de salud imaginable*) a 100 (*el mejor estado de salud imaginable*), además de la alternativa *muerte inmediata*, que podía coincidir o no con el extremo inferior del intervalo, así como situarse por encima de alguno de los estados sometidos a valoración, en cuyo caso se habría de interpretar que el sujeto consideraba dicho estado “peor que la muerte”. Entre los cinco estados

¹⁰ El mayor tamaño inicial de los subgrupos de la muestra 2 perseguía que, una vez aplicados los criterios de exclusión, el número de observaciones por cada nivel/dimensión no fuese muy inferior a 120.

puntuados en la escala se encontraba uno de los siete seleccionados para la obtención de la función de utilidad multiatributo objeto del presente estudio.

En el caso de la muestra 2, sendos “termómetros”, aunque esta vez delimitados por *el peor nivel posible* (valor 0) y *el mejor nivel posible* (valor 100) sirvieron para que los entrevistados colocasen en él los niveles intermedios correspondientes a cada una de las dos dimensiones que les tocó en suerte evaluar.

Teniendo en cuenta que, tanto las utilidades de los estados multiatributo como las utilidades uniatributo se medirían en otra parte de la encuesta a través de un método basado en elecciones (el *probability lottery equivalence* descrito en la sección 4.1.2), la puntuación de estados y niveles en la escala visual tenía como principal objetivo el familiarizar a los sujetos con el instrumento de clasificación y el modo en que éste describía los problemas de salud.

La parte tercera de la entrevista es la que incluía el método elegido para la obtención de utilidades, un escenario de “doble lotería” que denominamos *probability lottery equivalence* (PLE), que se repetía consecutivamente cinco veces en el cuestionario 1 (una por cada estado de salud multiatributo que le correspondía valorar al sujeto) y dos tres o cuatro veces (según el número de niveles de la dimensión evaluada), en dos ocasiones (una por cada dimensión que el encuestado hubiese de valorar) en el cuestionario 2. El orden de aparición de los estados dentro de esta sección en el cuestionario 1, así como el orden en el que se presentaron las dos dimensiones en el cuestionario 2 y la secuencia de aparición de los distintos niveles dentro de ellas, fueron rigurosamente aleatorios.

En esencia, el procedimiento PLE que aplicamos consistió, como se ha dicho en el apartado 4.1 de este informe, en encontrar el valor de p que consigue la indiferencia entre:

$$(SN, p, Muerte) \sim (SN, 0.5, Estado H) \quad (9)$$

o bien entre:

$$(Nivel 1, p, Nivel M) \sim (Nivel 1, 0.5, Nivel i) \quad (10)$$

según se tratase del primero o el segundo cuestionario.

Los valores iniciales de p se fijaban de manera aleatoria entre 0.55 y 0.95. A partir de aquí, sucesivas elecciones con distintos valores de $p > 0.5$ conducían a encontrar el valor de indiferencia.¹¹

La cuarta parte del cuestionario se destinó a recoger información sobre el estado de salud del propio encuestado, así como acerca de sus características personales (sexo, edad, nivel de estudios, situación laboral, etc.). Para acceder a la descripción y valoración del propio estado de salud se utilizaron varios instrumentos: el “auto-clasificador” EQ-5D, el cuestionario de salud SF-36 (versión 2) y una escala visual en forma de termómetro, sobre la que el sujeto debía valorar su estado de salud en relación con los extremos (*el mejor estado imaginable–el peor estado imaginable*), junto con la alternativa *muerte inmediata*. Para la muestra 2 también se dispuso de información acerca de la salud “objetiva” de los encuestados, mediante la realización de una serie de preguntas acerca de si la persona había sido diagnosticada o no de una serie de problemas de salud concretos.

4.3. Evaluación de la capacidad predictiva de la función estimada

Para evaluar la capacidad predictiva de la función multiatributo resultante del procedimiento descrito más arriba, se compararán los valores generados por la función con las utilidades observadas para un conjunto de estados SF-6D. Estos estados fueron los valorados mediante el método PLE al objeto de estimar el modelo de inferencia estadística (Abellán et al., 2008), salvo los siete incluidos en la aproximación algebraica, es decir, un total de 71 de los 78 directamente valorados.

Los criterios que se utilizarán para analizar la validez predictiva de la función serán la precisión (a través del cálculo del error medio cuadrático y el error absoluto medio), la mínima diferencia clínicamente significativa (calculando el porcentaje de estados con un error de predicción superior a 0.03 y 0.05), la sesgidez de las predicciones (mediante la prueba t) y el grado en el que el error de predicción está relacionado con la gravedad del estado de salud (a través del estadístico de Ljung Box).

¹¹ En el caso de la muestra 1, una pregunta inicial con $p=0.5$ permitía saber si el estado en cuestión se consideraba como “peor que la muerte”, en cuyo caso el sujeto elegiría la lotería de la izquierda y los valores de p que se presentaban en sucesivas elecciones eran siempre inferiores a 0.5, tratando de buscar con ello el mínimo valor de la probabilidad de “salud normal” que se estaba dispuesto a afrontar con tal de evitar un riesgo del 50% de sufrir el estado H .



Uno de los objetivos del estudio consiste en evaluar comparativamente el desempeño de los dos enfoques metodológicos (compuesto y descompuesto) que cabe seguir para estimar un algoritmo o tarifa de utilidades para un instrumento multiatributo. La comparación del enfoque basado en la teoría de utilidad multiatributo con la aproximación basada en inferencia estadística se realizará también en términos de capacidad predictiva. En este caso, para que la comparación sea lo más homogénea posible, ambos modelos se contrastarán con un panel de valores observados correspondientes a estados no intervinientes en ninguno de los protocolos de estimación. Ello requerirá reestimar el algoritmo basado en inferencia estadística, limitando a 48 el número de estados que se utilizan en la modelización, de manera que las predicciones de ambos modelos se contrasten con las utilidades directamente observadas de 23 estados de salud ($=78 - 48 - 7$) completamente ajenos tanto al análisis de regresión que conduce al algoritmo basado en inferencia como al procedimiento de obtención de los parámetros de la función de utilidad multiatributo.

Se utilizarán para esta evaluación comparativa los mismos tests de validez antes enunciados: error cuadrático medio, error absoluto medio, % de errores inferiores a 0.03 (0.05), t-test y prueba de Ljung-Box.

5. Resultados.

5.1. Participantes en el estudio

Se entrevistó a un total de 1230 individuos representativos por cuotas de edad y sexo de la población adulta, distribuidos en 2 muestras. La muestra 1 la componen 840 sujetos y la muestra 2 la forman 390 individuos. Las entrevistas se realizaron en el territorio de la Comunidad Autónoma de la Región de Murcia a lo largo del mes de marzo de 2008, para el caso de la muestra 1 (valoraciones multiatributo) y un año más tarde, en marzo de 2009, en el caso de la muestra 2 (valoraciones uniatributo). Los cuestionarios, asistidos por ordenador, fueron administrados en el domicilio de los encuestados por personal adiestrado por el equipo investigador. La duración media de las entrevistas fue de 20 minutos.

Del total de sujetos integrantes de la muestra 1, 13 fueron excluidos del análisis posterior por incurrir en alguna inconsistencia al valorar los estados de salud, de manera que un estado “lógicamente peor” recibía una valoración superior a un estado “lógicamente mejor”. De estos 13 sujetos, 3 incurrieron en alguna inconsistencia lógica al ordenar y puntuar los estados sobre la escala visual, mientras que en los otros 10 las inconsistencias afloraron en la respuestas al método de “doble lotería” (PLE). Otros 8 sujetos fueron excluidos de la muestra final debido a que, en la parte del cuestionario donde se elicitaban las utilidades con el PLE, se mostraron renuentes a aceptar un mínimo riesgo de muerte en al menos tres de los cinco estados que les correspondió valorar.¹² No hubo lugar a excluir ninguna encuesta incompleta, debido a que el diseño de la aplicación informática impedía el registro de la encuesta si no se cumplimentaba en su totalidad.

En definitiva, la muestra que se utiliza en el análisis subsiguiente está compuesta por 819 individuos. Por lo que respecta a la incidencia de los criterios de exclusión sobre los subgrupos en los que se dividió la muestra al objeto de “repartir” los estados de

¹² Los casos en los que el individuo no acepta riesgo alguno de morir con tal de evitar el estado crónico H dan lugar a que la utilidad del estado sea igual a 1. Esto ocurre cuando el sujeto escoge siempre la lotería en la que los *outcomes* son la salud normal ($p=0.5$) y el estado crónico H , por muy elevada que sea la probabilidad que se ofrece de recuperar la salud normal en la lotería que incluye la muerte como posible resultado. Si esto ocurre tres o más veces para un mismo sujeto, interpretamos que no es capaz de discriminar entre estados de salud según su gravedad y optamos por su exclusión.

salud, si bien resultó dispar, no hay ningún estado de salud para el cual se disponga de un número de observaciones inferior a 109, como se verá en el apartado siguiente.

En la muestra 2 se prescindió de las observaciones correspondientes a un total de 21 individuos. Uno de ellos incurrió en una inconsistencia lógica al puntuar en la escala visual los niveles de una de las dimensiones evaluadas (la puntuación del nivel 3 era inferior a la del nivel 2 en una de las dimensiones). Por haber incurrido en alguna inconsistencia lógica similar a la descrita, pero en las valoraciones del método de la doble lotería (PLE) se excluyó a otros 16 sujetos. Otro sujeto fue excluido porque no admitió ningún riesgo de muerte en las preguntas del método PLE (todas las utilidades uniatributo eran iguales a la unidad). Finalmente, tres sujetos fueron excluidos debido a que en las preguntas del método PLE no discriminaron entre niveles en ninguna de las dos dimensiones valoradas. En consecuencia, el tamaño de la muestra quedó finalmente en 369 sujetos y, aunque la aplicación de los criterios de exclusión afectó de modo dispar a los tres subgrupos de la muestra 2, se dispone de un mínimo de 121 observaciones para cada conjunto de utilidades uniatributo.

Tabla2. Descripción de las muestras

	<i>Muestra 1 (N=819)</i>	<i>Muestra 2 (N=369)</i>
% Mujeres	49,94	51,50
Edad media (desv. est.)	43,73 (16,78)	43,22 (16,83)
% Casado o pareja de hecho	58,49	55,56
% Separado, divorciado o viudo	7,45	5,42
% Estudios primarios	25,15	23,04
% Estudios secundarios	33,58	35,77
% Estudios superiores	28,45	23,85
% De 901 a 1.500 euros	16,48	15,99
% De 1.501 a 2000 euros	30,40	32,25
% De 2.001 a 3.000 euros	27,72	36,04
% Más de 3.000 euros	20,02	12,47
% Seguro médico privado	20,51	27,91
Salud autopercibida (EQ-5D)		
% 11111	62,64	74,80
% 11121	13,06	7,86
% 11112	4,15	5,15
Salud autopercibida (SF-6D/SF-36)		
% 111111	3,30	7,86
% 111112	4,40	5,96
% 111122	6,35	4,34

En la Tabla 2 se presentan las características de ambas muestras una vez aplicados los criterios de exclusión que, teniendo en cuenta la baja tasa de excluidos (4% para las dos muestras consideradas de forma conjunta), apenas difieren de las de las muestras originales.

Cabe destacar que la distribución de encuestados por características sociodemográficas básicas como la edad, el sexo y el estado civil, la distribución por niveles educativos y por intervalos de ingreso disponible en el hogar, es muy parecida en las dos muestras utilizadas. Si, como han destacado estudios previos como, por ejemplo, Dolan y Roberts (2002) o Kharroubi et al. (2007), las preferencias individuales sobre salud varían en función de características sociodemográficas y laborales, esta homogeneidad en la composición de las muestras avala la opción de combinar elementos de información de las dos muestras, como hacemos en este trabajo.

Las discrepancias más relevantes se observan en los indicadores de salud autopercebida, independientemente del instrumento utilizado para medir la salud. El porcentaje de encuestados que considera que está en una situación de salud perfecta es claramente inferior en la primera de las muestras.

5.2. Utilidades de los estados de salud SF-6D y utilidades “uni-atributo”

5.2.1. Utilidades multiatributo

En la Tabla 3 se ofrecen los estadísticos descriptivos de las utilidades de los 7 estados de salud SF-6D obtenidas mediante el método de “doble lotería”.

Tabla 3. Estadísticos descriptivos para los estados SF-6D

Estado	N	Media	Mediana	SD	P25	P75
115111	118	0,649	0,660	0,122	0,560	0,760
111115	114	0,855	0,900	0,197	0,800	0,940
411111	117	0,895	0,900	0,100	0,800	1,000
111411	119	0,803	0,780	0,105	0,740	0,880
111131	120	0,878	0,900	0,110	0,800	0,960
641111	115	0,672	0,700	0,264	0,540	0,900
645655	116	-0,515	-0,600	0,426	-0,900	-0,350

Cada estado fue valorado una media de 117 sujetos, con un mínimo de 114 y un máximo de 120. En cinco de los siete estados las medianas son superiores a las medias, lo que refleja una cierta asimetría en los datos agregados por estado. Sólo los estados

641111 y 111115, además del *pits state* fueron considerados peores que la muerte (utilidad < 0) por alguno de los sujetos encuestados. Por otra parte, en los casos en que resulta posible ordenar lógicamente los estados (el 645655 respecto de los seis restantes y el 641111 respecto del 411111), tanto las medias como las medianas verifican un criterio de consistencia lógica.

5.2.2. Utilidades uniatributo

En la Tabla 4 se describen las utilidades uniatributo correspondientes a los distintos niveles de las seis dimensiones del SF-6D, obtenidas mediante el método PLE (*probability lottery equivalence*), correspondiendo los valores de utilidad 1 y 0 al mejor y peor nivel, respectivamente, dentro de cada dimensión.

Tabla 4. Valores medios (desviaciones estándar) de las utilidades uniatributo

Dimensión	N	Nivel 2	Nivel 3	Nivel 4	Nivel 5
1. Funcionamiento físico	121	0,854 (0,105)	0,647 (0,129)	0,439 (0,153)	0,330 (0,136)
2. Limitaciones del rol	122	0,690 (0,171)	0,458 (0,159)	---	---
3. Funcionamiento social	126	0,826 (0,127)	0,620 (0,135)	0,384 (0,154)	---
4. Dolor	122	0,803 (0,115)	0,587 (0,139)	0,443 (0,138)	0,301 (0,135)
5. Salud mental	121	0,802 (0,104)	0,562 (0,133)	0,347 (0,165)	---
6. Vitalidad	126	0,834 (0,119)	0,629 (0,128)	0,404 (0,141)	---

La pérdida de utilidad derivada de una desviación de salud perfecta es, en todos los casos, creciente en la gravedad de la desviación. Asimismo, la dimensión en la que las desviaciones de salud perfecta resultan en pérdidas de utilidad más elevadas es 'limitaciones de rol'. En concreto, destaca la relevante pérdida de utilidad resultante de pasar de salud perfecta a tener algunas limitaciones en el trabajo por molestias físicas (nivel dos) en esta dimensión, una pérdida superior a la que se registra, para el paso del nivel 1 al nivel 2, en cualquier otra dimensión. Asimismo, el paso de los niveles tres a cuatro (nivel máximo) en esta dimensión se salda con la reducción de utilidad media más elevada de cuantas pueden calcularse a partir del cuadro anterior. La utilidad media uniatributo se reduce en 0,46 puntos en una escala 0 a 1, esto es, casi a la mitad, al pasar al nivel de gravedad máximo.

5.3. Análisis

5.3.1. Obtención de los parámetros de escalamiento

A diferencia de Torrance et al. (1996), las muestras a partir de las cuales hemos obtenido las valoraciones uniatributo que se resumen en la Tabla 4 y las utilidades multiatributo descritas en la Tabla 3 son distintas, con lo que la obtención de los parámetros de escalamiento, así como la formulación de la función de (des)utilidad no se realizará a nivel individual, sino a partir de los valores medios. Esta forma de proceder es, en nuestra opinión, válida habida cuenta del elevado tamaño de las muestras utilizadas, de la similitud en características sociodemográficas básicas entre ellas y de que el objeto del análisis es la obtención de tarifas medias y no el estudio de efectos distribucionales.¹³

En consecuencia, utilizando las expresiones (7) y (8) y las medias de los siete estados de salud multiatributo de la Tabla 3 identificamos los parámetros de escalamiento c y c_j , para los seis valores del parámetro j . Tal y como explican Feeny et al. (2002), c_j mide la importancia relativa de la dimensión de salud j en la determinación de la utilidad multiatributo, mientras que el parámetro c es una medida del grado de interacción en preferencias entre las distintas dimensiones de salud. Un valor negativo de este parámetro es indicativo de complementariedad en las preferencias sobre distintas dimensiones de salud, mientras que un valor positivo es evidencia de sustituibilidad entre las preferencias; el valor cero implica la ausencia de interacción entre dimensiones (modelo aditivo-lineal).

Las valoraciones de los estados de salud esquina 'puros' 115111 y 111115 permiten, a partir de la expresión (7), identificar directamente el valor de c_3 y c_6 . En concreto, el valor de c_j se corresponde con la desutilidad media del estado esquina 'puro' definido para la dimensión j , medida en la escala *all worst*-buena salud.¹⁴

¹³ Como se advierte en McCabe et al. (2005b), las encuestas diseñadas con el fin de generar una función de utilidad multiatributo poblacional no precisan de un conjunto completo de valoraciones para cada uno de los sujetos encuestados. En el apéndice de su artículo proponen, de hecho, un diseño de encuesta similar al que se ha utilizado en el presente estudio.

¹⁴ Las utilidades medias de la Tabla 3 están expresadas en la escala muerte-buena salud (u^*). Para determinar los valores de los parámetros de escalamiento, dichas utilidades han de convertirse a desutilidades ($\bar{u}^* = 1 - u^*$) y, posteriormente, ser transformadas a la escala *all worst*-buena salud, para lo cual es preciso dividir las por la desutilidad asignada al estado 645655: $\bar{u} = \bar{u}^* / 1,515$.

Un razonamiento similar permite identificar c_1 , c_4 y c_5 a partir de la desutilidad media de los estados 'cuasi-esquina' 411111, 111411 y 111131, respectivamente. En este caso se utilizan las desutilidades uniatributo para extrapolar hasta la esquina 'pura' y así calcular la desutilidad media multiatributo que se hubiera observado en caso de haberse valorado en la muestra el correspondiente estado de salud esquina 'puro'.

Finalmente, los valores de c_2 y c se identifican a partir del sistema de ecuaciones formado por la expresión (7) valorada en el estado 'doble-esquina' 641111 y la expresión (8). En este sentido, y a diferencia de Torrance et al. (1996), no analizamos la posibilidad de que las preferencias definidas sobre las dos dimensiones que definen el estado 'doble-esquina' estén relacionadas. Procedemos así por dos motivos.

En primer lugar, la corrección de la posible dependencia entre las preferencias asociadas a las dimensiones 'funcionamiento físico' y 'limitaciones de rol' requiere, como han destacado Stevens et al. (2007) al analizar el procedimiento utilizado en Torrance et al. (1996) para tal fin, de una serie de supuestos no contrastables acerca de cómo se relacionan los distintos niveles de severidad de las dos dimensiones y cómo se extrapolan utilidades. En segundo lugar, el análisis de las dependencias en preferencias asociadas a dimensiones de salud debería extenderse, tanto en Torrance et al. (1996) como en Stevens et al. (2007) y en el presente documento, a otras dimensiones de salud con interrelaciones lógicas entre sus niveles de gravedad, y no sólo a las consideradas en la definición del estado de salud 'doble-esquina'. Lo más razonable e inocuo para el análisis es, por tanto, no considerar la posible dependencia entre las dimensiones que definen el estado de salud 'doble-esquina' o cualquier otro conjunto de dos o más dimensiones de salud.

La Tabla 5 presenta los valores obtenidos para los parámetros de escalamiento. De acuerdo con estos valores, las dimensiones de salud 'funcionamiento social' y 'dolor' determinan en mayor medida, y con un peso relativo casi idéntico, la (des)utilidad de un estado de salud. En el extremo opuesto se encuentran las dimensiones 'limitaciones de rol' y 'vitalidad', con un peso relativo un 60% inferior al correspondiente a las dimensiones 'funcionamiento social' y 'dolor'. De las dos dimensiones restantes, 'salud mental' muestra un peso relativo en las preferencias asociadas a estados de salud un 50% superior al correspondiente a la dimensión 'funcionamiento físico'. Así, la

dimensión 'salud mental' ocupa el tercer puesto en importancia relativa, con una ponderación cerca de un 30% inferior a la de las dos dimensiones con mayor peso.

Tabla 5. Valores de los parámetros de escalamiento

Dimensión	Parámetro	Valor
1. Funcionamiento físico	c_1	0,124
2. Limitaciones del rol	c_2	0,093
3. Funcionamiento social	c_3	0,232
4. Dolor	c_4	0,234
5. Salud mental	c_5	0,184
6. Vitalidad	c_6	0,096
Interacción	c	0,000

El otro resultado relevante de la Tabla 5 es el valor de c . El valor obtenido, cero, es particularmente informativo, ya que indica ausencia de interacción en preferencias entre las dimensiones que definen un estado de salud. Esto es, la función de desutilidad en (7) puede reescribirse como el siguiente modelo lineal:

$$\bar{u} = \frac{1}{c} \left[\prod_{j=1}^6 (1 + c c_j \bar{u}_j) - 1 \right] = \sum_{j=1}^6 c_j \bar{u}_j \quad (12)$$

Conviene subrayar este hallazgo, por cuanto se aparta de los resultados de estudios previos. En las funciones multiatributo estimadas con anterioridad, el valor de c ha resultado ser, invariablemente, inferior a cero y muy próximo a -1: -0.97 en Torrance et al., (1982, 1996); -0.991 en Feeny et al. (2002); -0.935 y -0.752 en McCabe et al. (2005b).¹⁵ En todos estos estudios, por tanto, la función de utilidad resultó ser multiplicativa y los atributos complementarios entre sí. En nuestro caso obtenemos una función aditiva lineal, lo que implica la total ausencia de interacciones entre las dimensiones.

5.3.2. Función de utilidad multiatributo para el SF-6D

La Tabla 6 resume el procedimiento de estimación utilizado y el resultado obtenido al presentar las desutilidades unatributo (\bar{u}_j), los parámetros de escalamiento (c , c_j) y la función de desutilidad resultante (\bar{u}), que está definida en la escala en la que el cero se identifica con el peor estado de salud descrito por el instrumento SF-6D (*all worst* o

¹⁵ Feeny et al. (2002) para la función multiatributo del *HUI:3*; los restantes para la función del *HUI:2*. McCabe et al. (2005b) obtienen dos funciones multiatributo a partir de la utilización de dos funciones de conversión de los valores de la escala visual a utilidades *standard gamble*: una potencial y otra cúbica.

pits state). Para disponer de desutilidades en la escala muerte-buena salud es preciso reescalar, tal y como se muestra en la parte inferior de la tabla, multiplicando por la desutilidad del estado *all worst* en la escala muerte-buena salud.

Tabla 6. Función de desutilidad multiatributo

Funciones de desutilidad uniatributo						
Nivel	Funcionam. físico	Limitaciones de rol	Funcionam. social	Dolor	Salud mental	Vitalidad
	\bar{u}_1	\bar{u}_2	\bar{u}_3	\bar{u}_4	\bar{u}_5	\bar{u}_6
1	0,000	0,000	0,000	0,000	0,000	0,000
2	0,146	0,310	0,174	0,197	0,198	0,166
3	0,353	0,542	0,380	0,413	0,438	0,371
4	0,561	1,000	0,616	0,557	0,653	0,596
5	0,670		1,000	0,699	1,000	1,000
6	1,000			1,000		

Parámetros de reescalamiento						
c	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆
0	0,124	0,093	0,232	0,234	0,184	0,096

Fórmula (escala 'all worst'-buena salud)

$$\bar{u} = \sum_{j=1}^6 c_j \bar{u}_j = 0.124 \bar{u}_1 + 0.093 \bar{u}_2 + 0.232 \bar{u}_3 + 0.234 \bar{u}_4 + 0.184 \bar{u}_5 + 0.096 \bar{u}_6$$

Fórmula (escala muerte-buena salud)

$$\bar{u}^* = \bar{u} \times 1.515 \quad (\text{o bien: } \bar{u}^* = \bar{u} / 0.660)$$

Tabla 7. Función de utilidad multiatributo

Funciones de utilidad uniatributo						
Nivel	Funcionam. físico	Limitaciones de rol	Funcionam. social	Dolor	Salud mental	Vitalidad
	u_1	u_2	u_3	u_4	u_5	u_6
1	1,000	1,000	1,000	1,000	1,000	1,000
2	0,854	0,690	0,826	0,803	0,802	0,834
3	0,647	0,458	0,620	0,587	0,562	0,629
4	0,439	0,000	0,384	0,443	0,347	0,404
5	0,330		0,000	0,301	0,000	0,000
6	0,000			0,000		

Parámetros de reescalamiento						
c	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆
0	0,124	0,093	0,232	0,234	0,184	0,096

Fórmula (escala muerte-buena salud)

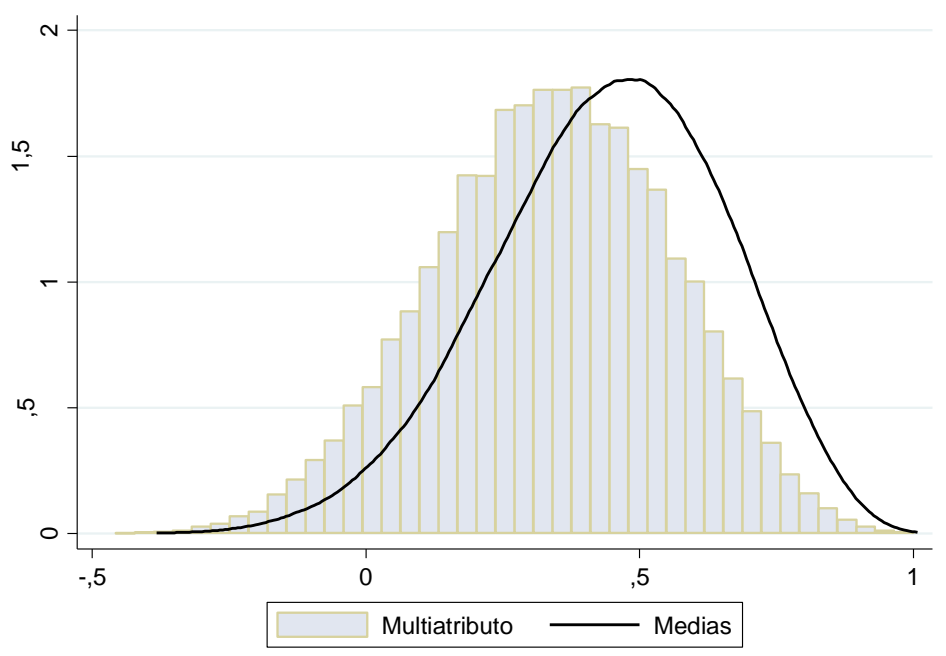
$$u^* = 1.515 \times 0.124 u_1 + 0.093 u_2 + 0.232 u_3 + 0.234 u_4 + 0.184 u_5 + 0.096 u_6 - 0.457$$

En la Tabla 7 se presenta la función multiatributo en términos de utilidad. Esta función (u^*) mide la utilidad media en la escala muerte-buena salud a partir de los parámetros c_j y las funciones de utilidad uniatributo (u_j). El coeficiente 1.515 en la función de utilidad se corresponde con la desutilidad media asignada al peor estado de salud (645655), en tanto que el término -0.457 que aparece al final de la función u^* se corresponde con el valor mínimo que predice la función (el valor que asigna al estado *all worst*) y resulta de:

$$1 - 1.515/\sum c_j$$

La función u^* permite predecir el nivel de utilidad medio de los 18.000 estados de salud que pueden definirse en el SF-6D. La distribución de estas utilidades se muestra en la Figura 2, en la que también se ha representado la distribución de los valores predichos por el modelo en medias estimado mediante MCO por el equipo investigador (Abellán et al., 2010), con el fin de ofrecer una referencia con la que comparar (más adelante se indaga en esta comparación).

Figura 2. Tarifas de utilidades predichas por la función multiatributo y el modelo de inferencia estadística de medias.



Al igual que se puso de manifiesto en Abellán et al. (2008, 2010) en relación con el algoritmo estimado mediante inferencia estadística, la función multiatributo de la Tabla 7 ‘rompe’ el umbral mínimo de la tarifa SF-6D estimada por Brazier et al. (2002)

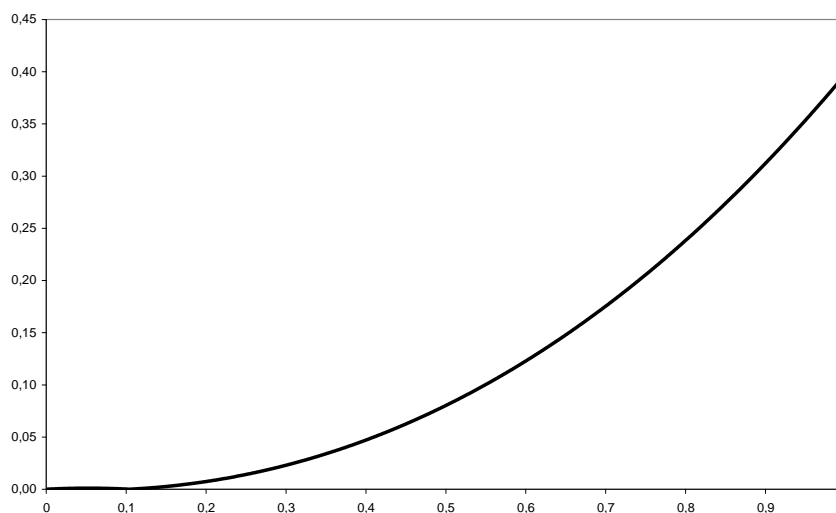
para el Reino Unido, extendiendo la cola izquierda de la distribución a niveles por debajo de cero. El valor mínimo que toma la función u^* es -0.457 , notablemente alejado del 0.296 que predice como valor más bajo la tarifa británica. Podemos concluir, pues, que el panel de utilidades que se desprende de la función multiatributo aquí presentada es inmune al denominado efecto ‘suelo’. La hipótesis que, a nuestro juicio, explica este resultado (Abellán et al., 2010) es que, el uso de un método de ‘lotería equivalente’ para medir las preferencias permite que las tarifas obtenidas a partir de nuestros datos no adolezcan del citado efecto, que da lugar a la falta de sensibilidad del instrumento para detectar ganancias de salud en individuos cuyo estado de salud de partida es malo.

Junto a este resultado cuyo interés se circunscribe al instrumento específico para el que se ha estimado la función multiatributo, el SF-6D, el segundo hallazgo de relevancia en nuestro estudio es, como se ha señalado más arriba, la obtención de una forma funcional aditiva lineal para \bar{u} , consecuencia del hecho de que la suma de los parámetros de escalamiento (c_j) es prácticamente igual a la unidad y, por tanto, el valor de c que hace que se verifique la expresión (8) es cero. No obstante, resulta oportuno introducir un matiz a este respecto, pues los análisis que se han realizado indican que la conclusión relativa a la linealidad de la función de utilidad multiatributo no es robusta.

Como se acaba de recordar, el valor del parámetro c en la función de utilidad multiatributo es aquel para el que se cumple la expresión (8). No obstante, la no linealidad de la expresión (8) en el parámetro c hace que puedan existir distintos valores de c para los que el grado de cumplimiento de la expresión (8) es prácticamente coincidente. Para analizar la verosimilitud de esta afirmación la Figura 3 representa la diferencia entre el lado derecho e izquierdo de la expresión (8) en función de los valores del parámetro c . La variable representada es una medida del grado de cumplimiento de la expresión (8) para distintos valores del parámetro c .

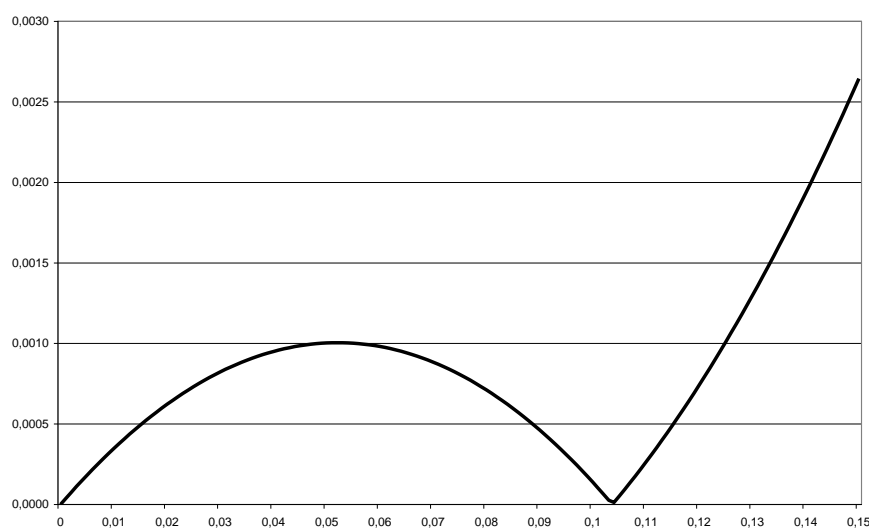
Esta figura pone de manifiesto que el grado de cumplimiento de la expresión (8) se reduce a medida que se incrementa el valor de c y que hay valores reducidos de c para los que el grado de cumplimiento es muy elevado o incluso coincidente.

Figura 3. Función a optimizar para determinar c y valor de c .



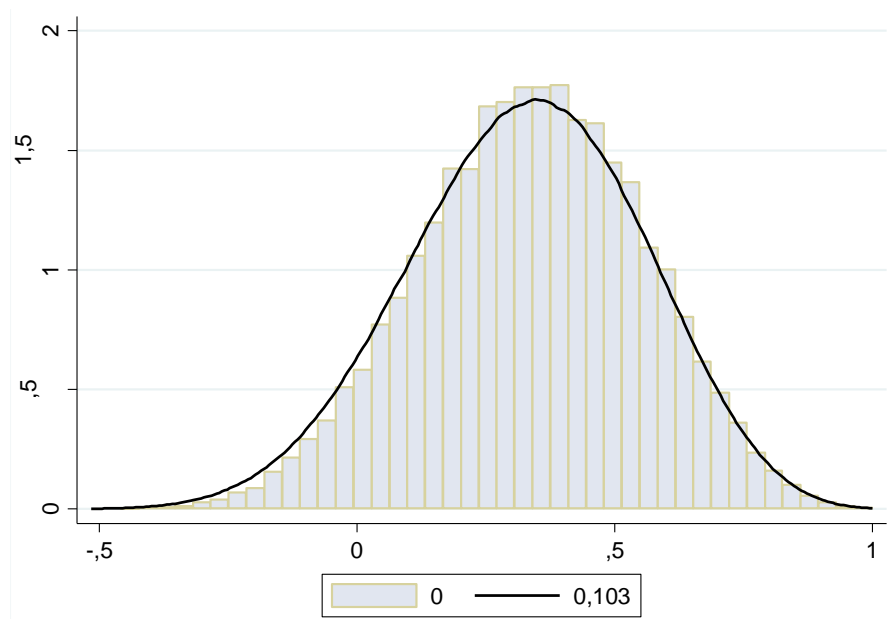
Para profundizar en este análisis, la Figura 4 presenta la misma información que la figura anterior pero centrándose en valores reducidos del parámetro c . En concreto, consideramos valores inferiores o iguales a 0,15 para ilustrar la no linealidad de la función objetivo en el parámetro de interés. Así, valores de c como 0,103 o 0,104 tienen valores de la función objetivo de 0,00001 y 0,00003 que, siendo superiores al valor puntual de cero que alcanza la función objetivo cuando el parámetro de interés toma ese mismo valor, son indicadores de que especificaciones no lineales de la función de (des)utilidad gozan de un grado de verosimilitud prácticamente igual a la del modelo lineal.

Figura 4. Función a optimizar para determinar c y valores reducidos de c .



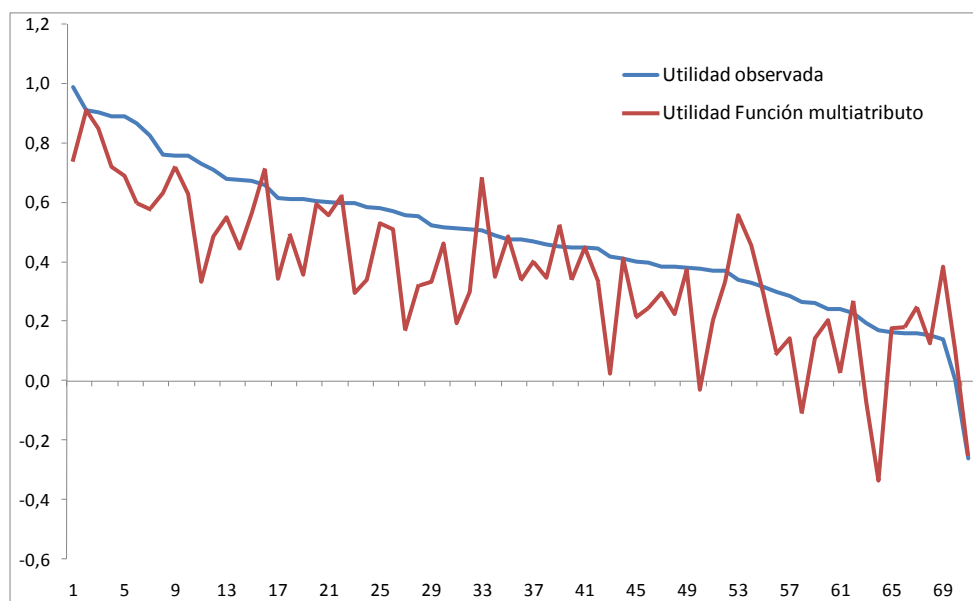
En todo caso, conviene aclarar que, tomando como valor de c cualquiera de los dos anteriormente señalados, obtenemos funciones multiatributo que predicen paneles de utilidades muy similares al que resulta del escenario de referencia en el que $c = 0$. Para ilustrar esta semejanza, en la Figura 5 se comparan las distribuciones resultantes de los modelos multiatributo en los que c toma los valores 0 y 0.103 respectivamente.

Figura 5. Tarifa multiatributo para distintos valores del parámetro c .



A la hora de juzgar la bondad de un modelo construido para generar las utilidades asociadas a un instrumento como el SF-6D, es habitual recurrir al criterio de capacidad predictiva, lo que exige comparar la utilidad que el modelo (la función multiatributo, en este caso) predice para un determinado estado de salud, con la utilidad directamente medida u observada. Con el fin de analizar la capacidad predictiva de la función de utilidad multiatributo, dicha función u^* se ha empleado para predecir la utilidad media de los estados de salud cuyas utilidades fueron objeto de medición directa a través del método PLE en la muestra 1, pero que no han sido tenidos en cuenta para la identificación de los parámetros de escalamiento y la posterior obtención de la función multiatributo. Es decir, utilizaremos en este análisis un total de 71 estados de salud: los que formaban parte de la selección de 78 que se empleó para modelizar las utilidades SF-6D a través del enfoque ‘compuesto’, a excepción de los 7 estados de la Tabla 1 que han servido de base para el enfoque ‘descompuesto’.¹⁶

¹⁶ Ver Tabla 14 en la página 58 de Abellán et al. (2008).

Figura 6. Utilidades medias observadas y predicciones del modelo multiatributo.

Por consiguiente, comparamos las utilidades que la función multiatributo predice para esos 71 estados de salud valorados en la muestra 1 con la utilidad media directamente medida para dichos estados a través del método PLE. El resultado se muestra gráficamente en la Figura 6, en la que los estados se han ordenado de manera decreciente según la utilidad directamente observada. Nuestra función multiatributo para el SF-6D presenta un error absoluto medio (EAM) de 0.150 y un error medio cuadrático (ECM) de 0.190. Un 16.7% de las predicciones se alejan menos de 0.03 del valor observado y un 25.0% de las desviaciones absolutas están por debajo de 0.05. El error medio resulta ser significativamente distinto de cero ($t = 10.76$), lo que revela que las utilidades que predice la función están sesgadas. Finalmente, el resultado del test de Ljung-Box pone de manifiesto que los errores de predicción no están relacionados con la gravedad del estado de salud ($LB = 30.98$).

Los resultados de estos tests de validez predictiva son muy similares a los que obtuvieron McCabe et al. (2005b) en la función multiatributo para el HUI:2 estimada según el protocolo seguido por Torrance et al. (1996). Su ejercicio de validación se basó en la comparación de las utilidades predichas por la función con las observadas en una muestra independiente para 65 estados de salud HUI:2 (frente a los 71 que hemos utilizado en nuestro estudio). En su caso el ECM y el EAM resultaron ser ligeramente superiores a los nuestros (0.206 y 0.182, respectivamente), y el porcentaje de estados con un error de predicción por debajo de los umbrales 0.03 y 0.05

sustancialmente inferior al obtenido con nuestra aproximación (7.7% en ambos casos). McCabe et al. (2005b) también obtuvieron un panel de utilidades sesgado ($t = -3.57$) en el que los errores de predicción no se relacionaron con la gravedad del estado de salud ($LB = 5.242$).

5.3.3. Comparación de la función multiatributo con el algoritmo SF-6D obtenido mediante inferencia estadística.

Para comparar los resultados del enfoque algebraico o ‘descompuesto’ con los derivados de aplicar técnicas de inferencia estadística (enfoque ‘compuesto’), hemos elegido, de entre los distintos algoritmos presentados en Abellán et al. (2008, 2010) el modelo que utiliza valores promedios como variable dependiente (modelo ‘de medias’ o ‘en medias’), que es el que suele ofrecer la mejor capacidad predictiva de entre los modelos de inferencia estadística utilizados en la estimación de tarifas.¹⁷

En la Figura 2 de la página 41 se observa claramente cómo el rango de valores en el que se mueven las dos tarifas representadas en la figura es muy similar, si bien la tarifa multiatributo presenta valores inferiores, en general, a los de la tarifa obtenida a partir del modelo de inferencia estadística. Como se ha señalado más arriba, ambas tarifas son inmunes al denominado efecto ‘suelo’. El resultado que añadimos a los ya contenidos en el estudio precedente es que el modelo multiatributo es más sensible que el de inferencia estadística a la hora de detectar mejorías a partir de estados de salud muy graves.

La linealidad obtenida para la función de utilidad multiatributo permite comparar directamente los coeficientes estimados utilizando el modelo de inferencia estadística con los niveles de utilidad asociados a cada nivel y dimensión de salud en el modelo multiatributo. Así, el ‘coeficiente’ asociado al nivel de gravedad k en la dimensión j en el modelo multiatributo (la desutilidad que conlleva dicho nivel en tal dimensión) se obtiene simplemente multiplicando el factor c_j por la desutilidad unatributo (\bar{u}_j) de dicha dimensión en el nivel k y convirtiendo el resultado a la escala muerte-buena

¹⁷ Por otra parte, la elección del modelo de medias parece congruente con el hecho de que la función multiatributo haya sido construida, igualmente, a partir de los valores medios de los estados SF-6D seleccionados y las (des)utilidades medias de las funciones ‘unatributo’.

salud.¹⁸ La columna ‘Función multiatributo’ de la Tabla 8 presenta estos coeficientes, junto con los del modelo de medias estimado por MCO a partir de la selección de 78 estados SF-6D (Abellán et al., 2010).

**Tabla 8. Modelización de estados de salud SF-6D en desutilidad.
Función multiatributo y modelo MCO en medias.**

	<i>Función multiatributo</i>	<i>Modelo de medias MCO</i>
FF2	0,027	0,015
FF3	0,066	0,034
FF4	0,105	0,090
FF5	0,125	0,111
FF6	0,187	0,338
LR2	0,044	0,014
LR3	0,076	0,038
LR4	0,141	0,070
FS2	0,061	0,037
FS3	0,133	0,060
FS4	0,217	0,203
FS5	0,351	0,208
DO2	0,070	0,018
DO3	0,146	0,034
DO4	0,197	0,198
DO5	0,248	0,202
DO6	0,354	0,318
SM2	0,055	0,066
SM3	0,122	0,078
SM4	0,182	0,096
SM5	0,279	0,224
VI2	0,024	0,058
VI3	0,054	0,121
VI4	0,086	0,157
VI5	0,145	0,199
Ajustado R2	-	0,967

FF: Funcionamiento físico; LR: Limitaciones del rol; FS: Funcionamiento social; DO: Dolor; SM: Salud mental; VI: Vitalidad.

¹⁸ Esto se consigue multiplicando el resultado del producto $c_j \bar{u}_j$ por la desutilidad del estado *all worst* en dicha escala, esto es, 1.515.

El modelo multiatributo predice desutilidades mayores para la mayoría de desviaciones de salud perfecta consideradas en el SF-6D. La excepción más notable es la dimensión 'vitalidad' (VI), para la que el modelo multiatributo predice utilidades superiores a las del modelo de inferencia estadística.

Las diferencias cuantitativamente más notables entre los dos conjuntos de estimaciones se localizan en las dimensiones 'salud mental' (SM), 'funcionamiento social' (FS), 'limitaciones del rol' (LR) y en los dos niveles más reducidos de la dimensión 'dolor' (DO). La magnitud de estas diferencias es más que notable. Así, los coeficientes multiatributo para los dos niveles más elevados de la dimensión LR doblan a los coeficientes estimados utilizando el modelo de medias. El coeficiente multiatributo asociado al nivel 2 de esta misma dimensión es tres veces mayor que el obtenido con el modelo de medias. Por su parte, los coeficientes asociados a los niveles 2 y 3 de la dimensión DO son, aproximadamente, cuatro veces mayores que los obtenidos con el modelo MCO.

En la sección anterior analizamos la capacidad predictiva del modelo multiatributo, obteniendo un EAM de 0.151 para los 71 estados cuyas utilidades habían sido directamente medidas en la muestra 1 y no habían contribuido a la formulación de la función de utilidad. La comparación de estos resultados con los del modelo de medias estimado mediante MCO arroja un balance claramente favorable a este último, pues el EAM se reduce a casi la mitad (0.081) y el porcentaje de errores de predicción inferiores a 0.1 es del 70,5% (35.2% en el modelo multiatributo) y las desviaciones absolutas por debajo de 0.01 representan un 11.7% (apenas un 6.5% en el multiatributo). Este balance tan favorable al modelo de inferencia estadística en términos de capacidad predictiva no debería extrañar, habida cuenta de que las utilidades observadas de los 78 estados que sirven para testar la validez predictiva constituyeron, a su vez, los inputs de la estimación del citado modelo de medias, mientras que los únicos 7 estados intervinientes en la obtención de la función multiatributo han sido excluidos en el cálculo de los errores de predicción (si se incluyen estos 7 el EAM se reduce a 0.126).

Así pues, para que el análisis comparativo entre ambos enfoques sea lo más objetivo posible, hemos reestimado el modelo de inferencia estadística en medias a partir de una muestra restringida que incluye solamente los 48 estados de salud resultantes de la ortogonalización¹⁹ (en lugar de los 78 de la estimación original).

**Tabla 9. Modelización de estados de salud SF-6D en desutilidad.
Modelo MCO en medias 'restringido'**

	Modelo de medias 'restringido' (N=48)
FF2	0,039
FF3	0,050
FF4	0,134
FF5	0,143
FF6	0,291
LR2	0,036
LR3	0,055
LR4	0,078
FS2	0,013
FS3	0,044
FS4	0,166
FS5	0,170
DO23	0,006*
DO4	0,178
DO5	0,239
DO6	0,347
SM2	0,056
SM34	0,082
SM5	0,216
VI2	0,053
VI3	0,116
VI45	0,215
Ajustado R2	0,971

FF: Funcionamiento físico; LR: Limitaciones del rol; FS: Funcionamiento social; DO: Dolor; SM: Salud mental; VI: Vitalidad. Algunas variables se han agregado (DO23, SM34, VI45) para evitar inconsistencias en los coeficientes. * coeficiente no significativo al 10%.

¹⁹ En realidad, el procedimiento de ortogonalización generó un subconjunto de 49 estados de salud, pero entre ellos se encontraba el estado *full health* (111111), cuya utilidad es fija e igual a la unidad.

La Tabla 9 presenta los coeficientes estimados utilizando el modelo de inferencia estadística en la submuestra resultante de la ortogonalización (48 estados). Con el fin de evitar inconsistencias en la predicción de las utilidades, se han agregado algunas variables indicadoras de gravedad.²⁰ Los coeficientes del modelo de medias restringido exceden a los del modelo no restringido (ver Tabla 8) en las dimensiones FF, LR y (DO, siendo inferiores en las dimensiones FS y VI. Las discrepancias, en aquellos niveles en los que es posible la comparación, pueden calificarse de reducidas, ya que no exceden del 25% del valor obtenido en el modelo no restringido en 11 de las 19 comparaciones posibles ni del 50% en 16 coeficientes.

Una vez obtenidos los nuevos coeficientes, analizamos la capacidad predictiva del modelo multiatributo y de este modelo en medias 'restringido' en los 23 estados de salud no utilizados por ninguno de los dos modelos (esto es, los 30 restantes hasta los 78 que se utilizaron en la estimación en medias original menos los 7 estados de la Tabla 1 empleados en la aproximación multiatributo). El resultado de este análisis comparativo se resume en la Figura 7, en la que los estados de salud se han ordenado por desutilidad (observada) creciente.

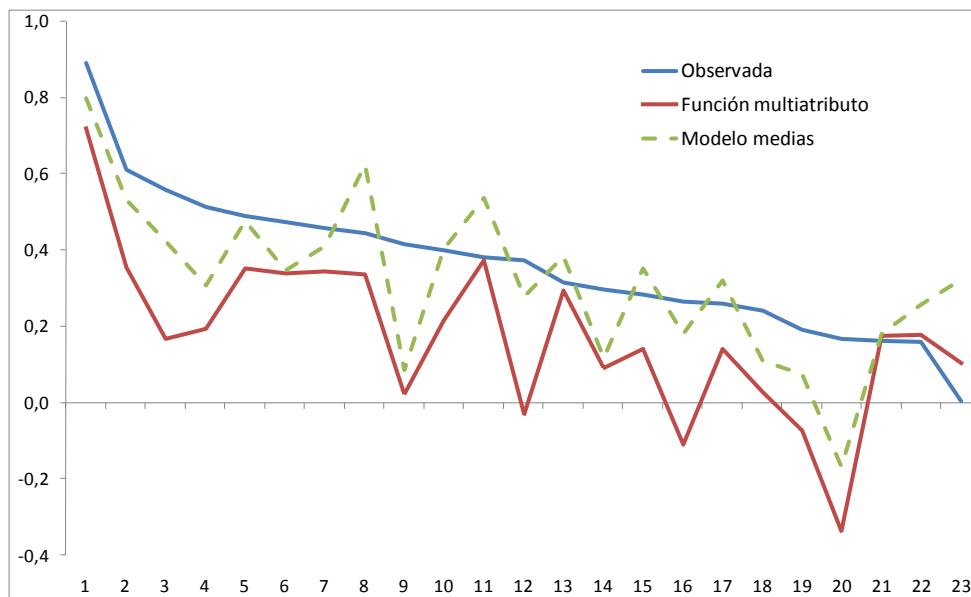
En la figura se aprecia claramente cómo el modelo econométrico (de medias) se ajusta mejor a las utilidades medias obtenidas en la muestra que el modelo algebraico multiatributo. Los tests de validez predictiva confirman esta impresión visual, pues el EAM del modelo multiatributo es 0.200, mientras que el correspondiente al modelo de medias es 0.128. Aunque la diferencia en términos de capacidad predictiva se ve atenuada en relación con la que se observaba al utilizar el modelo de medias estimado con la totalidad de observaciones (N=78), el EAM de este modelo 'restringido' sigue siendo más de un 30% inferior al del modelo multiatributo.²¹ Conclusión similar se deriva de comparar los errores medios cuadráticos: 0.243 en la función multiatributo y 0.159 en el modelo de inferencia estadística. El porcentaje de predicciones alejado

²⁰ Las inconsistencias a las que hacemos referencia aparecen cuando el coeficiente asociado a un nivel de gravedad de una dimensión es mayor en valor absoluto que el coeficiente asociado al nivel de gravedad inmediatamente posterior. Esta regla, no obstante, presenta algunas excepciones en el caso del SF-6D (ver Abellán et al., 2008, 2010).

²¹ En cualquier caso, ambas cifras son notablemente superiores a los EAM que se obtienen cuando la capacidad predictiva de los dos modelos se analiza utilizando los 78 estados valorados directamente en la muestra 1: 0.126 y 0.090, respectivamente.

menos de 3 ó 5 centésimas de las utilidades observadas es, sin embargo, prácticamente el mismo en ambos modelos.

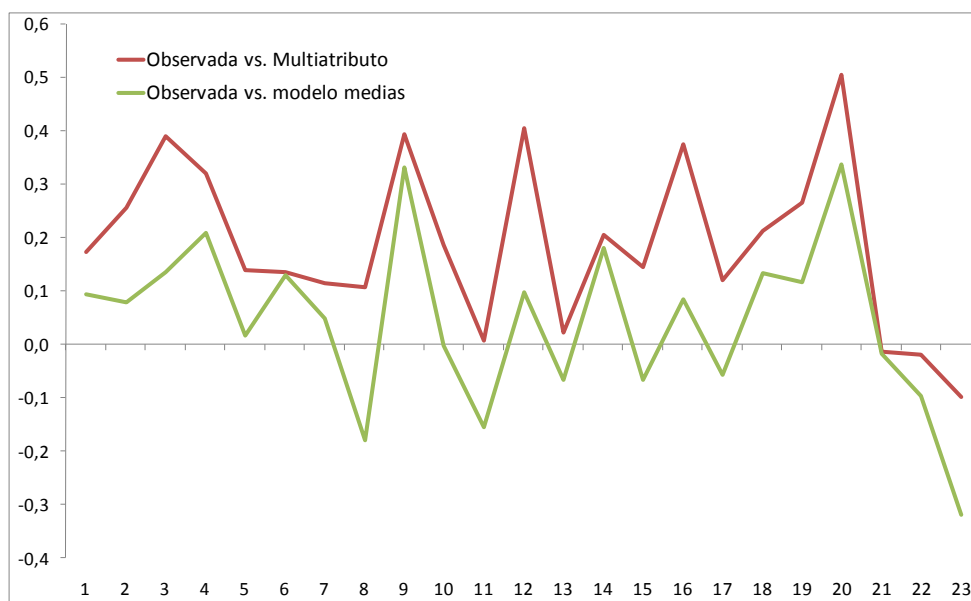
Figura 7. Utilidades medias observadas y predichas.



Ambos conjuntos de utilidades están sesgados, como queda de manifiesto en los valores obtenidos en la prueba t (10.76 en la función multiatributo; 7.01 en el modelo de medias 'restringido'). No obstante, como se observa en la Figura 7, las utilidades (desutilidades) medias obtenidas con el modelo multiatributo son sistemáticamente inferiores (superiores) a las predichas por el modelo de inferencia estadística (y también, con la excepción de los estados más graves, a las directamente obtenidas en la muestra).

La Figura 8 permite comprobar cómo el error absoluto por estado registra un valor medio y una variabilidad inferiores en el modelo de efectos aleatorios que en el modelo multiatributo. Sin embargo, en ninguno de los dos casos se aprecia un patrón de comportamiento de los errores de predicción en función de la gravedad del estado de salud. Esto queda confirmado con el resultado del test de Ljung Box, que también indica no significatividad en el modelo de inferencia estadística (LB = 5.10). De igual modo, tampoco se evidencia un patrón en las diferencias entre la utilidad media medida con el modelo multiatributo y con el modelo de inferencia estadística que pueda asociarse a la gravedad del estado.

Figura 8. Error absoluto medio por estado de salud.



Este resultado favorable a la superioridad predictiva del modelo de inferencia estadística es similar al que obtienen Stevens et al. (2007) al comparar el modelo multiatributo y el modelo de inferencia estadística de efectos aleatorios utilizando el instrumento HUI:2. No obstante, a diferencia de lo encontrado en nuestro análisis, en su estudio detectan una cierta correlación entre nivel de gravedad del estado de salud y discrepancia entre las predicciones de los modelos. En concreto, mientras que las diferencias entre las predicciones de los dos modelos son muy reducidas para estados leves de salud, aquéllas se incrementan de forma notable con la gravedad del estado analizado, siendo nuevamente el modelo multiatributo el que predice desutilidades medias más elevadas.

Para concluir con este análisis comparado de validez predictiva, es oportuno señalar que utilizando valores alternativos para el parámetro de interacción c (0.103 ó 0.104) se obtiene un EAM muy similar al del escenario de referencia en el que c vale cero. Esto es, la superioridad del modelo de inferencia estadística sí es una conclusión robusta de este estudio. Esta superioridad del modelo de inferencia estadística sugiere la ausencia de interacciones entre las dimensiones de un estado de salud en las preferencias de la población; una conclusión que, tal y como destacan en Stevens et al. (2007), contradice las previamente obtenidas en Dolan (1997) y Brazier et al. (2002) al estimar tarifas EQ-5D y SF-6D para el Reino Unido, respectivamente.

Siendo la capacidad predictiva un criterio fundamental a la hora de juzgar la idoneidad de la función de utilidad (o del algoritmo de cálculo de las utilidades) de un instrumento multiatributo como el SF-6D, conviene no obstante advertir de que en ningún caso puede el error de predicción erigirse en criterio único de valoración. La conclusión de nuestro estudio sobre el SF-6D, como la de Stevens et al. (2007) para el caso del HUI:2, es que el enfoque de inferencia estadística permite estimar un modelo que predice mejor las utilidades observadas que la función basada en el enfoque multiatributo. Esto conduciría a recomendar el uso del enfoque ‘compuesto’ frente al ‘descompuesto’ siempre que se pretendiese generar un panel de utilidades para un instrumento de calidad de vida relacionada con la salud como los mencionados.

Sin embargo, carece de sentido generar una prescripción de este tipo sin tomar en consideración el hecho de que los protocolos de investigación que dan sustento a una y otra aproximación metodológica son diferentes, lo que puede dar lugar a que uno y otro enfoque lleven aparejados distintos requerimientos en tiempo y otros recursos. Como se señaló en la sección 2 de este informe, la principal ventaja a priori del método ‘descompuesto’ radica en que minimiza el número de estados de salud que es necesario valorar de manera directa para luego poder realizar predicciones *out-of-sample*.

Esto se pone claramente de manifiesto en nuestro estudio, y debería ser tenido muy en cuenta, en nuestra opinión, como criterio complementario a la hora de enjuiciar la validez de los modelos, así como en el momento de optar por un enfoque metodológico o por el otro. Nuestra estimación basada en inferencia estadística del algoritmo del SF-6D exigió la valoración directa de un total de 78 estados de salud. La obtención de la función de utilidad basada en la teoría multiatributo para el mismo instrumento precisó de la valoración de 7 estados de salud SF-6D (los estados ‘esquina’ y similares, además del *all worst*) más las 19 utilidades ‘uniatributo’ correspondientes a los niveles intermedios de cada dimensión: 4 para los atributos con 6 niveles (FF y DO) 3 para los de 5 niveles (FS, SM y VI) y 2 para la dimensión LR, que tiene 4 niveles. Es decir, teniendo en cuenta que el método de elicitación fue el mismo en ambos casos (la ‘lotería equivalente’ PLE), el enfoque de inferencia estadística requirió del triple de tareas de valoración (78 frente a 26) que la aproximación algebraica. El resultado final

es que el modelo de inferencia tiene una mayor capacidad predictiva que el modelo multiatributo (EAM un 30% inferior), pero al 'coste' de triplicar el número de tareas de valoración y, por tanto, los requerimientos de recursos (tiempo y otros).

No queremos con esto decir que el enfoque multiatributo deba ser la opción preferida a la hora de abordar la estimación de una función de utilidad o un algoritmo de cálculo para un instrumento de medida de la CVRS. Pretendemos, por el contrario, advertir acerca del hecho de que la comparación 'cruda' del desempeño de los modelos alternativos en términos de capacidad predictiva está obviando posibles diferencias en el coste de obtención de los inputs que permiten construir los modelos. Teniendo en cuenta que en este tipo de estudios no es ajeno al clásico conflicto asignativo entre recursos escasos y necesidades (de información) virtualmente ilimitadas, lo que conduce a la inevitabilidad de afrontar *trade-off* (p. ej. número de estados de salud valorados frente a observaciones individuales por estado), las consideraciones relativas al coste no son cuestión baladí. Los costes de obtención de la información que sirve de base a los modelos pueden ser (en nuestro estudio claramente lo son) inferiores en el caso del enfoque multiatributo, por lo que esta aproximación metodológica puede en determinadas circunstancias resultar una alternativa coste-efectiva. Decimos 'puede resultar' porque asegurar si ésta es o no una opción metodológica coste-efectiva requeriría determinar en última instancia cuál es umbral mínimo admisible para los errores de predicción, tarea ésta que trasciende los objetivos del presente estudio.

6. Conclusiones.

Resumimos aquí los principales resultados del estudio, comenzando por los que tienen que ver con la obtención de la función de utilidad multiatributo para el SF-6D para, a continuación, glosar las conclusiones obtenidas al comparar estos resultados con los que se derivaron de la estimación mediante técnicas de inferencia estadística del algoritmo para este mismo instrumento y que se presentaron en Abellán et al. (2008).

6.1. Función de utilidad multiatributo para el SF-6D

El principal objetivo de esta investigación era la obtención de una función de utilidad para el instrumento SF-6D mediante el enfoque de la teoría de la utilidad multiatributo, también conocido como enfoque ‘descompuesto’. Los únicos precedentes similares a nuestro estudio, hasta donde llega nuestro conocimiento, son los trabajos de Torrance et al. (1982, 1996), Feeny et al. (2002) y McCabe et al. (2005b), todos ellos aplicados sobre otro instrumento de calidad de vida relacionada con la salud multiatributo, el *Health Utilities Index (HUI:2, HUI:3)*.

Nuestro estudio se diferencia de los anteriores en varios aspectos que, a nuestro juicio, representan una mejora respecto del proceder metodológico seguido en los citados trabajos. En primer lugar, hemos aplicado un método basado en elecciones para la obtención de las utilidades (desutilidades) que han servido de base a la formulación de la función multiatributo. Concretamente, se ha recurrido a una variante de ‘doble lotería’ conocida como *probability lottery equivalence (PLE)*, tanto para la medición de las utilidades de los estados de salud SF-6D (estados ‘esquina’, ‘cuasi-esquina’ y ‘doble esquina’), como para derivar las funciones de utilidad de cada uno de los atributos del instrumento considerados de forma aislada (funciones ‘uniatributo’). Por el contrario, los estudios precedentes utilizaron un método de medición de preferencias introspectivo, la escala visual analógica (EVA), junto con la estimación de una función de conversión (*mapping function*) para transformar las valoraciones de la EVA en utilidades *standard gamble*. Stevens et al. (2007) reconocen que el uso de datos procedentes de la EVA está sujeto a múltiples sesgos, lo que sin duda tiene un impacto significativo en la función de utilidad. Por un lado, la desutilidad de todos los estados multiatributo (estados ‘esquina’ o similares) se valora de manera sucesiva en la misma

escala, con lo que la propagación del error aumenta la probabilidad de que la suma de dichas desutilidades (que, posteriormente, se identificarán con las c_j) sumen más de la unidad. Esto condiciona, en última instancia, el hecho de que la función multiatributo sea multiplicativa. Por otro lado, al medir las desutilidades uniatributo en una única escala para cada dimensión, los valores obtenidos pueden estar afectados por el número de niveles en cada atributo, más que por el verdadero valor de tales condiciones de salud.

Una segunda característica de nuestro trabajo que, a nuestro juicio, otorga a sus conclusiones mayor validez que a las de estudios previos tiene que ver con el diseño del estudio para la recogida de datos. Los estudios previos optaron por estimar las funciones de utilidad a escala individual (además de la función basada en los valores medios), lo que exigió que cada individuo se enfrentase a la totalidad de las tareas de valoración, con el consiguiente riesgo de provocar sesgos asociados a la fatiga de los encuestados y de incrementar los errores de valoración como consecuencia de un volumen de información difícilmente abarcable por los sujetos.²² En el caso de nuestro estudio, la decisión de sustituir la escala visual como método de elicitación por una técnica basada en elecciones –el *probability lottery equivalence* (PLE)– incrementaba los requerimientos en tiempo y esfuerzo cognitivo hasta un punto que consideramos incompatible con la opción de obtener la totalidad de las valoraciones un mismo sujeto.²³

En consecuencia, decidimos aplicar un diseño ‘entre muestras’, tal y como se describió en el epígrafe 4.2, lo que nos permitió, por una parte, minimizar los errores y sesgos asociados a un elevado número de tareas por encuestado y, por tanto, a una duración excesiva de la entrevista; y, por otra parte, maximizar el número de observaciones para cada estado de salud (o nivel de utilidad ‘uniatributo’) valorado. La validez de este

²² Por ejemplo, en el estudio de Torrance et al. (1996) con el HUI:2 los entrevistados se enfrentaron a ocho tareas de valoración en la escala visual, siete de ellas para evaluar los niveles intermedios (entre 2 y 3) de cada uno de los siete atributos del instrumento (escalas uniatributo) y la última para puntuar un total de 13 estados multiatributo más la muerte. Adicionalmente, los sujetos hubieron de realizar cuatro tareas de valoración con el *standard gamble* para otros tantos estados HUI:2 multiatributo. Los autores informan de una duración media de las entrevistas próxima a los 65 minutos.

²³ Como se ha señalado en los párrafos finales de la sección anterior, el número total de valoraciones mediante el método PLE necesario para derivar la función multiatributo ascendía a 26 (7 valoraciones de estados SF-6D y 19 valoraciones de niveles intermedios en las funciones ‘uniatributo’).



diseño metodológico se procuró *a priori* garantizando la homogeneidad de las muestras en términos de sexo y cuotas de edad y se analizó *a posteriori* comparando la distribución de las características socio-demográficas de las distintas muestras sin que se detectasen diferencias significativas entre ellas.

Por lo que atañe a la consecución del principal objetivo del estudio, se ha estimado una función de utilidad multiatributo para el SF-6D, mediante un enfoque algebraico o ‘descompuesto’, a partir de las preferencias de una muestra de población murciana, representativa por cuotas de edad y sexo de la población española. Esta función, a diferencia de las estimadas para el Reino Unido (Brazier et al., 1998, 2002; Brazier y Roberts, 2004), no adolece del denominado efecto ‘suelo’, pues el rango de utilidades que genera para el instrumento SF-6D se extiende por debajo de cero, siendo así sensible a la medición de estados de salud cuya gravedad los hace tan indeseables que se consideran peores que la muerte. En este sentido, pues, el resultado del estudio confirma los obtenidos en la estimación de la tarifa SF-6D mediante inferencia estadística por el mismo equipo investigador (Abellán et al., 2008, 2010).

La principal singularidad de la función multiatributo estimada, en comparación con las obtenidas previamente para el instrumento HUI (Torrance et al., 1996; Feeny et al., 2002; McCabe et al., 2005b) radica en el hecho de que el valor del parámetro de interacción entre preferencias por atributos c toma un valor cero, indicando así ausencia total de interacción; los precedentes citados, por el contrario, obtuvieron invariablemente valores de c próximos a -1 (lo que se interpreta como que las preferencias son complementarias en las distintas dimensiones). En otras palabras, nuestra evidencia (aunque no del todo robusta, como se explica en la sección anterior) es favorable al modelo aditivo-lineal, en oposición a la forma multiplicativa obtenida por los mencionados trabajos previos. Este resultado es interesante por cuanto la linealidad, esto es, la independencia entre niveles de gravedad de las distintas dimensiones, es un resultado (cuando no un supuesto) habitual en la aproximación a las funciones de utilidad de instrumentos multiatributo basada en inferencia estadística.

6.2. Comparación de la función multiatributo con las tarifas estimadas mediante inferencia estadística

El segundo objetivo de nuestro trabajo era comparar la función multiatributo derivada a partir del enfoque algebraico con los algoritmos estimados mediante técnicas de inferencia estadística (enfoque ‘compuesto’). En Abellán et al. (2010) se seleccionó como modelo ‘preferido’ el resultante de estimar por MCO los coeficientes del algoritmo SF-6D a partir de las utilidades medias observadas para el subconjunto de 78 estados objeto de valoración directa. Es con este modelo ‘de medias’ o ‘en medias’ con el que hemos comparado nuestra función multiatributo, lo que nos ha llevado a introducir algunos cambios en la estimación econométrica con el fin de comparar la validez de ambos modelos (en particular su capacidad predictiva) en términos homogéneos.

El rango de valores en el que mueven las dos tarifas (es decir, los dos conjuntos o paneles de utilidades predichos por cada una de ellas) es muy similar aunque, en general, la tarifa asociada a la función multiatributo ofrece valores inferiores a los de la tarifa generada con el modelo econométrico de medias. En cualquier caso, tal y como se señala más arriba, ambas tarifas parecen ser inmunes al efecto ‘suelo’ característico del algoritmo SF-6D británico. Las diferencias más notables entre la función multiatributo y el algoritmo del modelo de medias se localizan en las dimensiones ‘salud mental’, ‘funcionamiento social’ y ‘limitaciones del rol’, así como en los niveles más reducidos de la dimensión ‘dolor’. Todas las dimensiones contribuyen a que las utilidades predichas con la función multiatributo sean inferiores a las generadas con el modelo de inferencia estadística, con la única excepción de la dimensión ‘vitalidad’, para la que el primero predice utilidades superiores.

Si se compara la capacidad predictiva de ambos modelos utilizando la totalidad de estados SF-6D directamente valorados (78) como referencia con la que contrastar las predicciones, se obtienen conclusiones claramente favorables al modelo de inferencia estadística. El valor del error absoluto medio (EAM) del modelo de medias es casi la mitad del obtenido con la función multiatributo (0.081 frente a 0.150). Sin embargo, cuando estimamos un modelo en medias ‘restringido’ (sólo con 48 estados como inputs) y valoramos la capacidad predictiva de éste y del modelo multiatributo

utilizando los 23 estados no intervinientes en ninguna de las modelizaciones, las diferencias en capacidad predictiva se atenúan (0.128 frente a 0.199).

En cualquier caso, la superioridad del modelo de inferencia estadística en términos de capacidad predictiva es un resultado robusto, lo que sugiere la ausencia de interacciones entre las dimensiones de un estado de salud en las preferencias de la población. Esta ausencia de interacciones es, precisamente, la que se pone de manifiesto en la forma lineal aditiva que derivamos en nuestra función multiatributo, como consecuencia del valor cero del parámetro c .

No obstante, la menor validez predictiva del modelo multiatributo no debería ser el único criterio, aunque sí uno de los más importantes, con el que enjuiciar el resultado de este enfoque metodológico alternativo. Como se ha señalado en varias partes de este informe, la principal ventaja del método algebraico radica en el menor requerimiento de valoraciones directas de estados de salud (o niveles uniatributo) para estimar el algoritmo de cálculo de las utilidades. Esta ventaja persiste y es palmaria si comparamos las 26 valoraciones que han resultado necesarias para estimar nuestra función multiatributo con los 78 estados de salud que fue preciso valorar con el fin de estimar los modelos de inferencia estadística del estudio que precedió a éste.

6.3. Aplicaciones de la función de utilidad multiatributo

La función de utilidad multiatributo para el SF-6D que aquí presentamos, al igual que el algoritmo basado en inferencia estadística estimado en su día, resulta de utilidad, al menos, en dos ámbitos: por un lado en el marco de las decisiones relativas a la asignación de recursos sanitarios en los niveles *meso* y *macro*; por otro en el contexto del abordaje de un análisis coste-beneficio generalizado de la sanidad regional.

En el primer caso, la función multiatributo permite generar una tarifa SF-6D con la que realizar ejercicios de evaluación económica de tecnologías y programas sanitarios. Para ello únicamente sería necesario disponer de la caracterización de los estados de salud relevantes para la evaluación (pre y post intervención) mediante el cuestionario SF-36 (o, directamente, mediante el sistema descriptivo SF-6D), el perfil temporal de los resultados de tales intervenciones y los costes de éstas. Estos elementos permitirían calcular ratios coste-efectividad (coste/AVAC) que servirían para decidir si la intervención evaluada es o no coste-efectiva ('vale lo que cuesta').



En segundo lugar, la disponibilidad de una tarifa de utilidades que permite asignar un índice cardinal de preferencias al resultado de un instrumento de tan extendido uso como el cuestionario SF-36 (del que se deriva el SF-6D) supone un primer paso hacia el objetivo de acometer un evaluación global de la rentabilidad social del gasto sanitario en la Región de Murcia. Este objetivo, como se ha puesto de manifiesto en informes previos, requeriría continuar transitando por la senda de colaboración abierta entre la Universidad de Murcia y la Consejería de Sanidad, hace más de 3 años, con el fin de abordar nuevos estudios con los que indagar acerca del valor social (monetario) del AVAC, así como sobre la contribución de la sanidad pública regional a la mejora de la salud de la población murciana.

Referencias

Abellán Perpiñán JM, Sánchez Martínez FI, Martínez Pérez JE, Méndez Martínez I. (2008) "Estimación de un algoritmo para el cálculo de utilidades de estados de salud descritos mediante el SF-6D(SF-36)". Memoria final del proyecto *Indicadores de calidad de vida relacionada con la salud y su aplicación a la evaluación de tecnologías sanitarias*, desarrollado en el marco del Convenio de Colaboración entre la Consejería de Sanidad y la Universidad de Murcia, en materia de Economía de la Salud, - BORM Nº 23, 29/01/2007.

Abellán Perpiñán JM, Sánchez Martínez FI, Martínez Pérez JE, Méndez Martínez I. (2010) "Breaking the 'floor' of the SF-6D utility function. An application to Spanish data". Mimeo (disponible en <http://www.um.es/esyec/gtes/>).

Badia X, Roset M, Herdman M, Kind P. A Comparison of United Kingdom and Spanish General Population Time Trade-off Values for EQ-5D Health States. *Med Decision Making* 2001; 21: 7-16.

Bleichrodt H, Abellan-Perpiñan JM, Pinto JL, Mendez I. Resolving Inconsistencies in Utility Measurement under Risk: Tests of Generalizations of Expected Utility. *Management Science* 2007; 53: 469-482.

Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *New England Journal of Medicine* 1983; 308(22): 1330-1337.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; 21: 271-92.

Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Medical Care* 2004; 42(9): 851-9.

Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 health survey. *J Clin Epidemiol* 1998; 51: 1115-28.

Brooks RG, Jendfeg S, Lindgren B, Persson U, Björk S. EuroQol: health-related quality of life measurement. Results of the Swedish questionnaire exercise. *Health Policy* 1991; 18(1): 37-48.



Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? *Soc Sci Med* 2002; 54: 919-29.

Dolan P. "Modelling the relationship between description and valuation of health states". En Murray C., Salomon J., Mathers C., López A. (eds). *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva, World Health Organization. 2002.

Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997; 35: 1095-108.

Essink-Bot ML, Bonsel GJ, van der Maas PJ. Valuation of health states by the general public: Feasibility of standardised measurement procedure. *Soc Sci Med* 1990; 31: 1201-6.

EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199-208.

Feeny D, Furlong W, Torrance GW, Goldsmith C, Zenglong Z, Depauw S, Denton M, Boyle M. Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Medical Care* 2002; 40(2): 113-28.

Furlong W, Feeny D, Torrance GW, Goldsmith C, DePauw S, Boyle M, Denton M, Zenglong Z. *Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report*, McMaster University Centre for Health Economics and Policy Analysis Working Paper No. 98-11. 1998.

Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley. 1976.

Kharroubi SA, Brazier J, Roberts J, O'Hagan A. Modelling SF-6D health state preference data using a nonparametric Bayesian method. *J Health Econ* 2007; 26: 597-612.

Kind, P. *Measuring valuations for health states: Piloting the EuroQol questionnaire*, Discussion paper No. 76. York: University of York, Centre for Health Economics, 1990.

Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value in Health* 2008; 11: 295-303.



Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ* 2006; 15: 1121-32.

McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the Health Utilities Index Mark 2 descriptive system: results from a UK valuable survey. *Health Econ* 2005a; 14: 231-44.

McCabe C, Stevens K, Brazier J. Utility scores for the Health Utilities Index Mark 2: an empirical assessment of alternative mapping functions. *Medical Care* 2005b; 43(6): 627-35.

McCord M, de Neufville R. Lottery equivalents: Reduction of the certainty effect problem in utility assessment. *Management Science* 1986; 32(1): 56–60.

Nord E. EuroQol: health-related quality of life measurement. Valuations of health states by the general public in Norway. *Health Policy* 1991; 18(1): 25-36.

Pinto JL, Abellán-Perpiñán JM. Measuring the health of populations: the veil of ignorant approach. *Health Economics* 2005; 14: 69 – 82

Stevens K, McCabe C, Brazier J, Roberts J. Multi-attribute utility function or statistical inference models: A comparison of health state valuations models using the HUI2 health state classification system. *J Health Econom* 2007, 26: 992-1002.

Torrance GW, Boyle MH, Horwood SP. Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States. *Operations Research* 1982; 30(6): 1043-69.

Torrance GW, Feeny D, Furlong W, Barr R, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Medical Care* 1996; 34(7):702–722.

Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002; 11: 341-353.

Apéndice 1. Los cuestionarios

Presentación

Tanto el cuestionario orientado a obtener las utilidades de los estados multiatributo (cuestionario 1) como el diseñado al objeto de derivar las funciones de utilidad uniatributo (cuestionario 2) se iniciaban con una presentación. Tras ella, en el cuestionario 1 se describía al entrevistado el sistema de clasificación de estados de salud SF-6D, explicándole los seis atributos o dimensiones que éste incluye, así como los niveles (entre 4 y 6) que cada dimensión puede admitir y la asociación entre niveles y “gravedad” de los problemas de salud. A tal fin se diseñaron unas “tarjetas” que, a modo de ayuda visual, pretendían ayudar a una mejor comprensión y valoración de las descripciones basadas en el instrumento multiatributo.²⁴

En el cuestionario 2 se explicaba a los encuestados que, de entre las distintas dimensiones que componen el estado de salud de una persona, la entrevista se centraría únicamente en dos; las que hubiera correspondido evaluar en cada caso agrupadas en los siguientes pares (“Funcionamiento físico”–“Salud mental”, “Limitaciones del Rol”–“Dolor, y “Funcionamiento social”–“Vitalidad”).

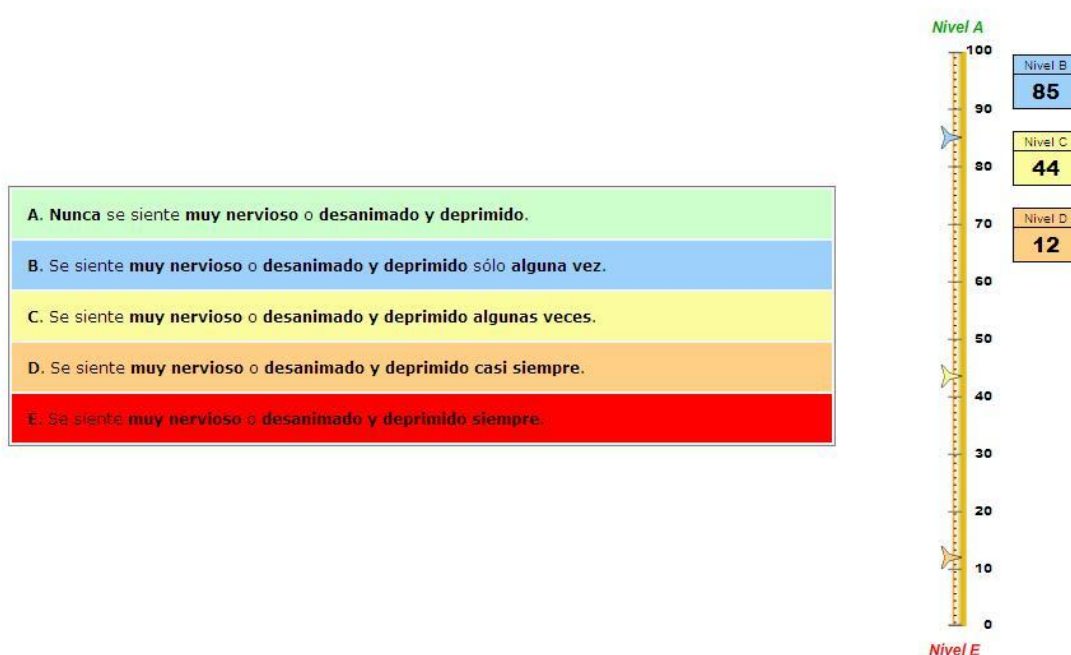
Puntuación de estados en la escala visual

Los sujetos que contestaron el cuestionario 1 hubieron de puntuar en una escala visual con forma de termómetro cinco estados de salud del sistema SF-6D, etiquetados anónimamente como V, W, X, Y y Z siendo los extremos inferior y superior de la escala *el peor estado de salud imaginable* y *el mejor estado de salud imaginable*, respectivamente, y en el que también ha de situar la opción *muerte inmediata*. Entre estos cinco estados se encontraba alguno de los siete que se han utilizado en este estudio basado en el enfoque algebraico o descompuesto.

²⁴ Información más detallada acerca del cuestionario 1 puede encontrarse en el Apéndice de la memoria “Estimación de un algoritmo para el cálculo de utilidades de estados de salud descritos mediante el SF-6D(SF-36)”, de Julio de 2008, correspondiente al proyecto *Indicadores de calidad de vida relacionada con la salud y su aplicación a la evaluación de tecnologías sanitarias*, desarrollado en el marco del Convenio de Colaboración suscrito con fecha 22 de diciembre de 2006 entre la Consejería de Sanidad y la Universidad de Murcia, en materia de Economía de la Salud, - BORM Nº 23, 29/01/2007.

Quienes respondieron al cuestionario 2 realizaron una tarea similar solo que, en este caso, los extremos del termómetro eran ocupados por el mejor y el peor nivel del atributo objeto de evaluación, y los encuestados debían situar sobre la escala los niveles intermedios. En la figura 1 se muestra una imagen de la aplicación informática en la que se puede observar un ejemplo del resultado final de esta tarea, que se repetía más adelante con el segundo de los atributos que le hubiera correspondido evaluar al entrevistado.

Figura 1



Obtención de medidas directas de utilidad (PLE)

La parte que contiene la aplicación del método de “doble lotería” elegido para obtener las utilidades –el *probability lottery equivalence* (PLE), que se expuso en la sección de Metodología–, se inicia en el cuestionario 2 del modo que se muestra más abajo en la Figura 2. Se repetía esta tarea para cada uno de los niveles intermedios que el sujeto había de evaluar dentro de cada una de las dos dimensiones. Tanto el orden de aparición de los dos atributos objeto de valoración en cada cuestionario como, dentro de ellos, la secuencia en la que aparecían los distintos niveles con el fin de ser evaluados fueron resultado de un proceso aleatorio:²⁵

²⁵ En el cuestionario 1 el método PLE se aplicó sucesivamente a los cinco estados valorados, de un modo similar al que aquí se describe para el cuestionario 2, si bien precedido de una pregunta inicial destinada

Figura 2

Imagine que su médico le informa de que es usted portador de una cierta enfermedad cuyo pronóstico es incierto. En concreto, en cincuenta de cada cien casos como el suyo, el paciente vive el resto de su vida sin sufrir ningún síntoma, esto es, en condiciones normales de salud. Sin embargo, en otros cincuenta casos de cada cien la enfermedad progresa provocando ciertos problemas que serán crónicos, es decir, que le acompañarán durante el resto de su vida. Concretamente, en caso de que la enfermedad avance, usted disfrutará de una salud normal en todos los aspectos salvo en el hecho siguiente:

D. Se siente muy nervioso o desanimado y deprimido casi siempre.

Su médico le explica, también, que existe un tratamiento que podría curarle completamente, pero dicho tratamiento es arriesgado. En ochenta de cada cien casos como el suyo el tratamiento es plenamente efectivo, pero en otros veinte casos de cada cien, los efectos adversos del tratamiento agravan el problema que pretendían resolver, de manera que durante el resto de su vida se enfrentaría a la siguiente situación (recuerde que el resto de aspectos de su salud serían normales):

E. Se siente muy nervioso o desanimado y deprimido siempre.

En consecuencia, si usted se somete al tratamiento tendrá un 80% de probabilidad de curarse y olvidarse completamente de la enfermedad y otro 20% de probabilidad de que su salud, en lo relativo al estado anímico se agrave al nivel E. Si, por el contrario, decide no someterse al tratamiento, se enfrenta a una probabilidad del 50% de vivir sin problemas de salud y a un 50% de probabilidad de sufrir los problemas descritos en el nivel D de manera crónica, es decir, durante el resto de su vida. Ante esta situación le pedimos, por favor, que responda:

¿Se sometería al tratamiento?

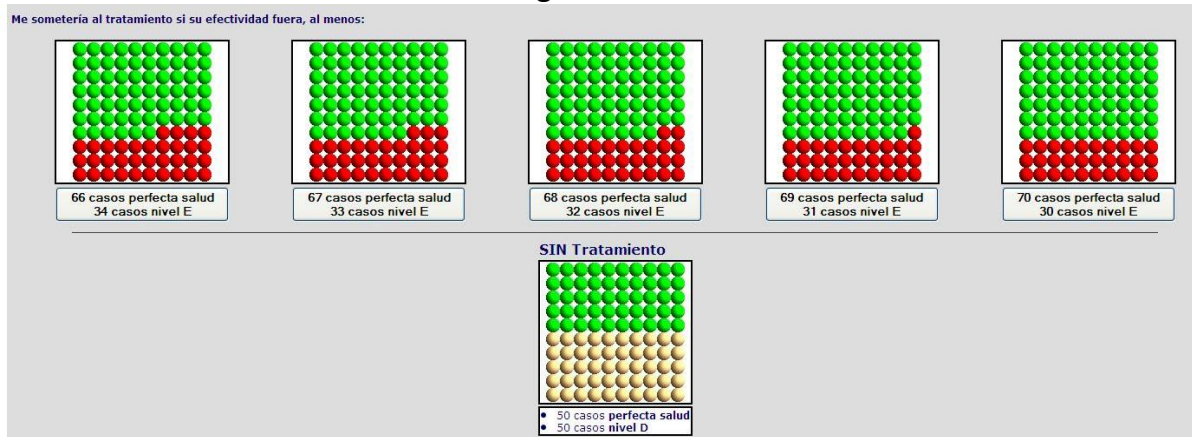


La probabilidad asociada a la salud normal que se ofrece como primer “estímulo” al sujeto (80% en la figura del ejemplo) es generada de manera aleatoria por la aplicación informática en el intervalo de 55 a 95. En función de cuál sea la respuesta del encuestado a esta pregunta el programa le “ofrecerá” una probabilidad menor –si elige la opción SÍ (*tratamiento*)– o una probabilidad mayor –si optó por la alternativa NO (*sin tratamiento*)–, manteniéndose constantes en el 50% las probabilidades de la salud normal y del estado crónico en la alternativa *SIN tratamiento*. Los valores que se van generando en las sucesivas preguntas siguen la secuencia basada en el procedimiento *PEST* que se explica al final de este apéndice, hasta acotar un intervalo del 5% en cuyo interior la respuesta del sujeto cambia. Supongamos, por ejemplo, que el individuo elige el tratamiento para una probabilidad del 65% de disfrutar de salud normal y lo rechaza para una probabilidad de salud normal del 60%. Llegado este

a identificar si el estado en cuestión era considerado por el sujeto “peor que la muerte”, pues el procedimiento de elecciones sucesivas que conduce al sujeto a revelar sus preferencias era diferente según se trate de un estado de salud preferido a la muerte o no. Para más detalles consultar el Apéndice de la memoria “Estimación de un algoritmo para el cálculo de utilidades de estados de salud descritos mediante el SF-6D(SF-36)”.

punto, se pide al sujeto que precise la probabilidad que haría que su decisión cambiase, planteándole una pregunta abierta como la de la Figura 3.

Figura 3



Autoevaluación del propio estado de salud

En ambos cuestionarios se incluyeron tres instrumentos de medición de la salud subjetivos: el “autoclasicador” EQ-5D, el cuestionario SF-36 (versión 2) y una escala visual para que los encuestados puntuasen su estado de salud actual. En el cuestionario 2 se añadieron preguntas acerca de la salud “objetiva” del encuestado, esto es, se le preguntó si se le había diagnosticado por parte de un médico una serie de patologías y problemas de salud concretos.

Así, el sujeto debía caracterizar su estado de salud en el momento de realizar la entrevista según las dimensiones y niveles del sistema EQ-5D (Figura 4).

Figura 4

<p>Movilidad</p> <p><input checked="" type="radio"/> No tengo problemas para caminar</p> <p><input type="radio"/> Tengo algunos problemas para caminar</p> <p><input type="radio"/> Tengo que estar en cama</p>	<p>Cuidado Personal</p> <p><input checked="" type="radio"/> No tengo problemas para realizar mi cuidado personal</p> <p><input type="radio"/> Tengo algunos problemas para lavarme o vestirme</p> <p><input type="radio"/> Soy incapaz de lavarme o vestirme por mí mismo</p>
<p>Actividades Cotidianas (ej. trabajar, estudiar, hacer las tareas domésticas, disfrutar del tiempo libre)</p> <p><input checked="" type="radio"/> No tengo problemas para realizar mis actividades cotidianas</p> <p><input type="radio"/> Tengo algunos problemas para realizar mis actividades cotidianas</p> <p><input type="radio"/> Soy incapaz de realizar mis actividades cotidianas</p>	<p>Dolor, Molestias</p> <p><input type="radio"/> No tengo dolor ni malestar</p> <p><input checked="" type="radio"/> Tengo dolor o malestar moderado</p> <p><input type="radio"/> Tengo mucho dolor o mucho malestar</p>
<p>Ansiedad o Depresión</p> <p><input type="radio"/> No estoy ansioso ni deprimido</p> <p><input checked="" type="radio"/> Estoy moderadamente ansioso o deprimido</p> <p><input type="radio"/> Estoy muy ansioso o deprimido</p>	

A continuación, los entrevistados debían cumplimentar la versión 2 del cuestionario de salud SF-36, que se administró dentro de la misma aplicación informática mediante un sistema de “botones-radio” (ver el extracto de la Figura 5, a modo de ejemplo). Esta

parte de la aplicación (también el autoclasificador EQ-5D) se programó de manera que resultase imposible avanzar en la encuesta si se había dejado sin responder alguna de las preguntas (o de los ítems que componen cada una de ellas). De este modo se garantizó un 100% de respuestas válidas en esta parte de la encuesta.

Figura 5

6. Durante las 4 últimas semanas, ¿hasta qué punto su salud física o los problemas emocionales han dificultado sus actividades sociales habituales con la familia, los amigos, los vecinos u otras personas?

Nada Un poco Regular Bastante Mucho

7. ¿Tuvo dolor en alguna parte del cuerpo durante las 4 últimas semanas?

No, ninguno Sí, muy poco Sí, un poco Sí, moderado Sí, mucho Sí, muchísimo

8. Durante las 4 últimas semanas, ¿hasta qué punto el dolor le ha dificultado su trabajo habitual (incluido el trabajo fuera de casa y las tareas domésticas)?

Nada Un poco Regular Bastante Mucho

9. Las preguntas que siguen se refieren a cómo se ha sentido y cómo le han ido las cosas durante las 4 últimas semanas. En cada pregunta responda lo que se parezca más a cómo se ha sentido usted. Durante las últimas 4 semanas ¿con qué frecuencia...

	Siempre	Casi siempre	Algunas veces	Sólo alguna vez	Nunca
se sintió lleno de vitalidad?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
estuvo muy nervioso?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
se sintió tan bajo de moral que nada podía animarle?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
se sintió calmado y tranquilo?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tuvo mucha energía?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
se sintió desanimado y deprimido?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
se sintió agotado?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Por último, se pedía a los encuestados que valorasen su propio estado de salud sobre una escala visual. En el caso del cuestionario 1 se utilizó para ello la misma escala que al inicio del cuestionario sirvió para valorar los cinco estados de salud evaluados y la muerte, “recuperada” a tal fin por la aplicación informática en esta última parte del cuestionario. Los sujetos que respondieron al cuestionario 2 situaron su propio estado de salud actual sobre una escala visual cuyos extremos eran *el peor estado de salud imaginable* y *el mejor estado de salud imaginable*; sobre esta misma escala se les pidió que situasen la *muerte*, toda vez que las escalas que habían cumplimentado previamente eran escalas de un solo atributo y no se disponía de su valoración subjetiva de la muerte en un intervalo con extremos como los descritos.

Características socio-demográficas

La parte final del cuestionario se dedicó a recoger información sobre características del entrevistado que posteriormente habría de utilizarse en el proceso de análisis de los datos, así como acerca de algunas actitudes personales. Por último se pedía a los entrevistados que indicasen el grado de dificultad que les había supuesto responder a las diferentes preguntas contenidas en la encuesta.

Procedimiento para la obtención de los valores de indiferencia (PEST) 26

En la parte de la encuesta en la que se obtienen las valoraciones mediante el procedimiento PLE, el valor de indiferencia se identificó siguiendo un sistema de obtención de respuestas secuencial como el sugerido por Luce²⁷, con el fin de evitar los sesgos asociados a los métodos basados en *matching*. El valor del parámetro con el que se pretende conseguir la indiferencia –en nuestro caso, la probabilidad asociada al resultado “salud normal” en la alternativa *Tratamiento*– va cambiando según la secuencia que más adelante se describe, en función de las elecciones que realiza el sujeto, hasta acotar el mínimo intervalo del 5%; llegados a este punto se pide al entrevistado que concrete (con precisión de 1 punto porcentual) la probabilidad que le hace ser indiferente entre recibir tratamiento o no.

Las normas que siguió el procedimiento en nuestro estudio fueron las siguientes:

1. El valor inicialmente propuesto se determina de manera aleatoria dentro del rango de valores posibles (del 55% al 95%).
2. Tras la primera respuesta del sujeto, la magnitud de la duración que se ofrece al sujeto aumenta (disminuye) en 10 puntos porcentuales (“paso inicial”).
3. Una segunda elección coincidente –por ejemplo, elige “NO (*Sin tratamiento*)” dos veces seguidas– da lugar al mismo incremento (disminución) en la probabilidad ofrecida.
4. A partir de la cuarta elección en un mismo sentido, el incremento (disminución) se duplica, siempre y cuando no se excedan los límites; en caso contrario, se fija una nueva probabilidad con un valor distante en una unidad de dicho límite.
5. Cada vez que el sujeto cambia su elección (pasa, por ejemplo, de elegir “Sí (*Tratamiento*)” a elegir “NO (*Sin Tratamiento*)”, el incremento (disminución) en la probabilidad se divide por dos.
6. La tercera elección en el mismo sentido puede dar lugar a un incremento (disminución) en la probabilidad ofrecida de igual o doble magnitud que el

²⁶ Acrónimo de *Parameter Estimation by Sequential Testing*.

²⁷ Luce RD. *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. New Jersey: Lawrence Erlbaum Associates, Inc. 2000.



anterior, según cuál haya sido el último cambio de elección. Si éste fue precedido de un aumento en la magnitud del incremento (disminución), ésta se mantiene; en caso contrario la magnitud de la variación se duplica.

7. Cada dos elecciones se ofrece al sujeto una distribución de probabilidades en la alternativa “*Sí (Tratamiento)*” totalmente aleatoria, sin relación alguna con la secuencia utilizada. Con ello se intenta que el sujeto no perciba que se le está “conduciendo” a lo largo de un proceso de convergencia.

Apéndice 2. Ejemplo de aplicación de la función de utilidad multiatributo a partir de las respuestas al cuestionario SF-36.

El primer paso necesario para el cálculo de utilidades consiste en traducir las respuestas al cuestionario SF-36 a las seis dimensiones del SF-6D. Para ello hay que seleccionar las preguntas 3a, 3b, 3j, 4c, 5b, 7, 8, 9e, 9b, 9f y 10 del SF-36 y traducirlas a los niveles del SF-6D. Por ejemplo, si a la pregunta 8 del SF-36 (“¿Hasta qué punto el dolor le ha dificultado su trabajo habitual?”) el sujeto responde “Bastante”, se le asignará el nivel 5 en la dimensión “Dolor” del SF-6D. Si, por ejemplo, a la pregunta 3a del SF-36 (“Su salud actual, ¿le limita para hacer ... esfuerzos intensos, tales como correr, levantar objetos pesados, o participar en deportes agotadores”) el sujeto responde “No, no me limita nada”, le corresponderá el nivel 1 en la dimensión “Funcionamiento físico”. Y así, sucesivamente.

Una vez hemos traducido las respuestas al SF-36 al sistema SF-6D, pasaremos a aplicar la función de utilidad que se presenta en la Tabla 7, sustituyendo u_1 , u_2 , etc. por la utilidad uniatributo correspondiente a los niveles de cada dimensión que caractericen el estado de salud del sujeto. Así, por ejemplo, si la situación de un individuo en un determinado momento del tiempo resultase descrita mediante el estado **231254**, calcularíamos el “peso de calidad de vida” o la utilidad del estado de salud de dicho sujeto como:

$$u^* = 1.515 \times 0.124 u_{1/2} + 0.093 u_{2/3} + 0.232 u_{3/1} + 0.234 u_{4/2} + 0.184 u_{5/5} + 0.096 u_{6/4} - 0.457$$

Donde $u_{1/2}$, $u_{2/3}$, $u_{3/1}$, $u_{4/2}$, $u_{5/5}$, $u_{6/4}$ aluden las utilidades uniatributo del nivel 2 de “funcionamiento físico”, el nivel 3 de “limitaciones del rol”, el nivel 1 de “funcionamiento social” el nivel 2 de “dolor”, el nivel 5 de “salud mental” y el nivel 4 de “vitalidad”, respectivamente. Sustituyendo los valores de la Tabla 7 tendríamos que la utilidad del estado 231254 es igual a:

$$u^* = 1.515 \times \left(\begin{array}{l} 0.124 \times 0.854 + 0.093 \times 0.458 + 0.232 \times 1.000 + \\ 0.234 \times 0.803 + 0.184 \times 0.000 + 0.096 \times 0.404 \end{array} \right) - 0.457 = \mathbf{0.461}$$



Alternativamente, se puede obtener la utilidad de dicho estado de salud restando de la unidad los correspondientes coeficientes de la Tabla 8, en la que el modelo se expresa en términos de desutilidad:

$$u^* = 1 - FF2 - LR3 - DO2 - SM5 - VI4$$

Donde FF2, LR3, DO2, SM5 y VI4 aluden los coeficientes del nivel 2 de “Funcionamiento físico”, el nivel 3 de “Limitaciones del rol”, el nivel 2 de “Dolor”, el nivel 5 de “Salud mental” y el nivel 4 de “Vitalidad”, respectivamente (la dimensión “Funcionamiento social” no supone pérdida alguna de utilidad, pues se encuentra en el nivel 1). Sustituyendo los valores de la columna ‘Función multiatributo’ de la Tabla 8 obtendríamos el mismo valor para el estado 231254:

$$u^* = 1 - 0.027 - 0.076 - 0.070 - 0.279 - 0.087 = \mathbf{0.461}.$$

El modelo de inferencia estadística (modelo en medias MCO de la Tabla 8) asigna a este mismo estado una utilidad significativamente más alta: 0.548.