UNIVERSIDAD DE MURCIA

FACULTAD DE ECONOMÍA Y EMPRESA

Avances metodológicos en el análisis de costes y resultados para la evaluación económica de intervenciones sanitarias

Dª. Silvia Garrido García

2014

UNIVERSIDAD DE MURCIA

Programa de Doctorado en Desarrollo y Cooperación Internacional

Avances metodológicos en el análisis de costes y resultados para la evaluación económica de intervenciones sanitarias

Memoria de investigación para optar al Grado de Doctor presentada por Silvia Garrido García, bajo la dirección del Dr. José María Abellán Perpiñán y del Dr. Fernando Ignacio Sánchez Martínez.

Fdo.: Silvia Garrido García

Facultad de Economía y Empresa

Murcia, 2014

A mis padres,

Rafael y Rosario

[If] for a second you turn back

(Si por un segundo retrocedes)

Oh no, be strong

(No lo hagas, sé fuerte)

[…]

Walk on, walk on

(Sigue adelante, sigue adelante)

What you've got they can't deny it

(Lo que tienes no lo pueden negar)

Can't sell it or buy it

(No lo pueden vender ni comprar)

Walk on, walk on

(Sigue adelante, sigue adelante)

Fragment of the song *Walk on* (U2),
included in the album *All that you can't leave behind*.

(Fragmento de la canción *Walk on* (U2),
incluida en el álbum *All that you can't leave behind*).

AGRADECIMIENTOS

Aprovecho estas líneas para mostrar mi agradecimiento a todas aquellas personas que, de una u otra forma, han contribuido a la realización de esta tesis:

A mis directores –José María Abellán y Fernando Sánchez– y a mis otros dos compañeros del Grupo de Trabajo en Economía de la Salud –Ildefonso Méndez y Jorge Martínez. A todos ellos les estoy infinitamente agradecida por haber confiado en mí desde el primer momento mucho más de lo que yo confío en mí misma; por haberme dado la oportunidad de descubrir el mundo de la Economía de la Salud, algo totalmente desconocido para mí antes de que yo decidiera hacer el doctorado hace cuatro años; por toda la atención y tiempo que me han dedicado; por todo lo que me han enseñado; y por la enorme paciencia y generosidad que han tenido conmigo.

Al resto de miembros del Departamento de Economía Aplicada, por su apoyo y amabilidad durante estos cuatro años.

A Renske Hoefman, Job van Exel y Werner Brouwer –investigadores del *Institute of Health Policy & Management* (iBMG), *Erasmus University Rotterdam*–, por su gran hospitalidad, generosidad y apoyo durante mis tres meses de estancia en dicho centro, así como por su ayuda inestimable para el diseño del estudio presentado en el Capítulo 3 de esta tesis. Igualmente, agradezco al resto de personal del iBMG, por haberme ayudado a sentir integrada desde el primer momento.

A mi familia, sobre todo a mis padres –Rafael y Rosario–, por haberme acompañado durante este largo proceso y por haberme permitido dedicar todo el tiempo necesario a la realización de este trabajo. Un agradecimiento especial se merecen mis sobrinos –Javier, Lola y José–, por haberme ayudado a desviar un poco mi atención de la tesis.

A mis amigas Toñi –mi querida compañera de trabajo en el CEBAS– y María, por su apoyo moral, oraciones y sabios consejos.

A la Universidad de Murcia, por el apoyo financiero que he recibido a través del programa de becas-contrato predoctorales 2010-2013. Por último, también manifiesto mi agradecimiento a la Dirección General de Planificación, Ordenación Sanitaria y Farmacéutica e Investigación de la Comunidad Autónoma de la Región de Murcia; a la Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia (proyecto 15375/PHCS/10); y al Ministerio de Economía y Competitividad (proyecto ECO2010-22041-CO2-02), por haber financiado los tres estudios que componen esta tesis.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who, in some way or another, have contributed to make this thesis possible:

To my supervisors –José María Abellán y Fernando Sánchez– and my other two colleagues of the Health Economics Research Group –Ildefonso Méndez y Jorge Martínez. I am infinitely grateful to all of them for having trusted me since the first moment, even more than I trust myself; for having provided me the opportunity to discover the world of Health Economics, which was something completely unknown for me four years ago, before I decided to pursue a PhD; for all the attention and time they have devoted to me; for all that they have taught me; and for their extraordinary patience and generosity towards me.

To the rest of members of the Department of Applied Economics, for their support and kindness during these four years.

To Renske Hoefman, Job van Exel and Werner Brouwer –researchers at the Institute of Health Policy & Management (iBMG), Erasmus University Rotterdam– for their great hospitality, generosity and support during my three-month stay at this centre, as well as for their invaluable help with the design of the study presented in Chapter 3 of this thesis. Likewise, I give thanks to the rest of the iBMG staff, for having helped me feel integrated since the first moment. *Dank je wel*!

To my family, above all to my parents –Rafael and Rosario–, for having accompanied me throughout this long process and for having allowed me to devote all the time necessary to this work. Special thanks are due to my nephews and niece –Javier, José and Lola–, for having helped me divert my attention a little away from the thesis.

To my friends Toñi –my dear workmate at the CEBAS– and María, for their moral support, prayers and wise advice.

To the University of Murcia, for the financial support that I have received through the pre-doctoral fellowship program 2010-2013. Finally, I am also very grateful to the Department of Health of the Region of Murcia; to the Seneca Foundation (grant 15375/PHCS/10); and to the Ministry of Economy and Competitiveness (grant ECO2010-22041-CO202), for having funded the three studies reported in this thesis.

TABLE OF CONTENTS

List of tables ………………………………………………………………. iii
List of figures ……………………………………………………………... iv
List of abbreviations ……………………………………………………… iv
Resumen general ………………………………………………………….. vii
General abstract ………………………………………………………….. xv

1 General introduction ……………………………………………………..... 1
1.1 The context surrounding the topics addressed in this thesis 3
1.1.1 The rationale for the economic evaluation of health care interventions 3
1.1.2. The basics of the economic evaluation of health care interventions 4
1.2 The relationship between health-related quality of life and life satisfaction 14
1.3 The monetary valuation of informal care 16
1.4 The effect of learning on the consistency of preferences 19

2 Analysing the simultaneous relationship between life satisfaction and health-related quality of life 23
Abstract 24
2.1 Introduction 25
2.2. Methods 27
2.2.1 Participants 27
2.2.2 The questionnaire 28
2.2.3 The model 32
2.3 Results 33
2.3.1 Background statistics 33
2.3.2 The unidirectional approach 36
2.3.3 The simultaneous equations model 40
2.4 Discussion 46
2.5 Appendix 2A 51
2.6 Appendix 2B 60

3 Monetary valuation of informal care based on carers' and non-carers' preferences 63
Abstract 64
3.1 Introduction 65
3.2 Methods 70
3.2.1 Participants 70
3.2.2 The questionnaire 71
3.2.3 Hypotheses 77
3.2.4 Data analysis 80
3.3 Results 82

3.3.1 Sample characteristics 82
3.3.2 Analysis of protest zeros 87
3.3.3 WTA values in reference to the hypothetical scenario 88
3.3.4 Comparison between carers' WTA values under the hypothetical and the actual caring situations........ 94
3.3.5 Regression analysis 97
3.4 Discussion 101
3.5 Appendix 3A 108
3.6 Appendix 3B 152

4 Testing the effect of learning on preference reversals 153
Abstract 154
4.1 Introduction 155
4.1.1 Taxonomy of preference reversals 155
4.1.2 Explanations for preference reversals 161
4.1.3 The effect of learning on preference reversals 163
4.1.4 Study objectives 167
4.2 Methods 167
4.2.1 Study design 167
4.2.2 Hypotheses 173
4.3 Results 175
4.3.1 Background statistics 175
4.3.2 Comparison between standard preference reversal and non-standard preference reversal at the aggregate level 180
4.3.3 The effects of repetition and feedback at the aggregate level 182
4.3.4 Analysis by groups 184
4.3.5 Analysis by starting point presented in the valuation tasks 188
4.3.6 Analysis by gender 191
4.4. Discussion........ 195
4.5 Appendix 4A 201
4.6 Appendix 4B 225

5 General conclusions 229
5.1 The simultaneous relationship between life satisfaction and health-related quality of life 231
5.1.1 Main results and discussion 232
5.1.2 Limitations and suggestions for future research 235
5.2. The monetary valuation of informal care 236
5.2.1 Main results and discussion 237
5.2.2 Limitations and suggestions for future research 239
5.3 The effect of learning on the consistency of preferences 240
5.3.1 Main results and discussion 241
5.3.2 Limitations and suggestions for future research 243

References 247

LIST OF TABLES

Table 2.1	Sample and Spanish population personal characteristics	34
Table 2.2	Life satisfaction and self-perceived health statistics	35
Table 2.3	Unidirectional approach. OLS standardised estimates	37
Table 2.4	Bivariate probit estimates of the correlation between the error terms of the two equations	40
Table 2.5	Simultaneous equations approach. 3SLS estimates	43
Table 2.6	Simultaneous equations approach by gender and age groups. 3SLS estimates	45
Table 2B.1	Dimensions and levels of the SF-6D descriptive system	60
Table 3.1	Overview of the questionnaires ………………………………………..	72
Table 3.2	Respondent's and Spanish population's characteristics	82
Table 3.3	Care-related characteristics according to age	86
Table 3.4	Classification of responses ……………………………………………….	88
Table 3.5	Tests of differences related to the protest zeros ………………………….	88
Table 3.6	WTA values , descriptive statistics (€/hour/day) ………………………...	89
Table 3.7	Test of differences regarding the three $WTA_{hypothetical}$ variables …………	89
Table 3.8	Logical consistency at the individual level ………………………………	91
Table 3.9	WTA_{worst} values by types of tasks (€/hour/day) …………………………	92
Table 3.10	WTA_{best} values by types of tasks (€/hour/day) …………………………	93
Table 3.11	Comparison between carers' $WTA_{general}$ and WTA_{own} values (€/hour/day)	94
Table 3.12	Comparison between carers' $WTA_{general}$ and WTA_{own} values at the individual level	96
Table 3.13	List of explanatory variables used in the regression analysis ……………	97
Table 3.14	OLS regressions for carers' WTA	98
Table 3.15	OLS regressions for non-carers' WTA	98
Table 4.1	Demographic characteristics and distribution of the sample	168
Table 4.2	Lotteries used in the experiment	169
Table 4.3	Structure of the two experimental sessions (example) …………………...	172
Table 4.4	CE values, summary statistics ……………………………………………	176
Table 4.5	Comparison between CE and expected values …………………………...	177
Table 4.6	Comparison between CE(P) and CE($) values …………………………..	178
Table 4.7	Switch of CE values ……………………………………………………...	178
Table 4.8	Choices …………………………………………………………………...	179
Table 4.9	Switch of choices ………………………………………………………...	179
Table 4.10	Classification of responses at the aggregate level………………………...	181
Table 4.11	Effects of repetition and feedback at the aggregate level ………………..	183
Table 4.12	Classification of response by groups …………………………………….	186
Table 4.13	Proportions of respondents (per group) who exhibited risk aversion in the choice and valuation tasks ……………………………………………	187
Table 4.14	Effects of repetition and feedback by groups …………………………….	188
Table 4.15	Classification of responses by stating point format of the valuation tasks ………………………………………………………………………	190
Table 4.16	Effects of repetition and feedback by format of the valuation tasks ……..	191

Table 4.17	Classification of responses by gender ……………………………………	193
Table 1.18	Proportions of men and women who exhibited risk aversion in the choice and valuation tasks ……………………………………………………….	194
Table 4.19	Effects of repetition and feedback by gender …………………………….	194

LIST OF FIGURES

Figure 2.1	Distribution of the SF-6D index	36
Figure 2.2	Distribution of the SWLS summary score	36
Figure 3.1	Description of the hypothetical scenario and wording of the $WTA_{general}$ question ………………………………………………………………..	75
Figure 3.2	Wording of the WTA_{own} question ……………………………………...	76
Figure 3.3	Difference between carers' $WTA_{general}$ and WTA_{own} values ……………	95
Figure 3B.1	WTA values	151
Figure 4B.1	Screenshot of a CE question	225
Figure 4B.2	Screenshot of a choice question	226
Figure 4B.3	Screenshot of the resolution of risk	227

LIST OF ABBREVIATIONS

BDM	Becker, DeGroot, Marschak
CA	Conjoint analysis
CBA	Cost-benefit analysis
CE	Certainty equivalent
CEA	Cost-effectiveness analysis
CI	Confidence interval
CarerQol-7D	Care-related quality of life 7 dimensions
CUA	Cost-utility analysis
CV	Contingent valuation
DCEs	Discrete choice experiments
DRM	Day Reconstruction Method
EQ-5D	EuroQol-5 dimension questionnaire
EU-27	European Union, 27 countries
EV	Expected value
HRQoL	Health-related quality of life
HUI3	Health Utilities Index Mark 3
IMSERSO	Instituto de Mayores y Servicios Sociales
IQR	Interquartile range
LR	Likelihood ratio

NHS	National Health Service
NICE	National Institute for Clinical Excellence
NSPR	Non-standard preference reversal
OLS	Ordinary least squares
PANAS	Positive and Negative Affect Schedule
PE	Probability equivalent / probability equivalence
PLE	Probability lottery equivalent / probability lottery equivalence
PPP	Purchasing power parity
PR	Preference reversals
PSS	Personal social services
RSE	Robust standard error
SD	Standard deviation
SE	Standard error
SF-6D	Short Form-6 dimension questionnaire
SF-36	Short Form-36 Health Survey
SG	Standard gamble
SPA	Second-price Vickrey auction
SPR	Standard preference reversal
StLPA	Second-to-last price auction
SWB	Subjective well-being
SWLS	Satisfaction With Life Scale
TTO	Time trade-off
VAS	Visual analogue scale
VE	Value equivalent / value equivalence
VLE	Value lottery equivalent / value lottery equivalence
vNM EUT	von Neuman-Morgenstern expected utility theory
WBV	Well-being valuation
WTA	Willingness to accept
WTA_{best}	WTA for one extra hour of care per day performing the most preferred task (in reference to the hypothetical scenario)
$WTA_{general}$	WTA for one additional hour of care per day in general terms (in reference to the hypothetical scenario)
$WTA_{hypothetical}$	This variable encompasses $WTA_{general}$, WTA_{worst} and WTA_{best}
WTA_{own}	WTA for one additional hour of care per day under the actual caring situation
WTA_{worst}	WTA for one extra hour of care per day performing the least preferred task (in reference to the hypothetical scenario)
WTP	Willingness to pay
2SLS	Two-stage least squares
3SLS	Three-stage least squares

RESUMEN GENERAL

El objetivo global de esta tesis es examinar nuevos enfoques metodológicos referentes al análisis de costes y resultados para la evaluación económica de intervenciones sanitarias. En concreto, estos nuevos enfoques se aplican en el contexto de la estimación de resultados relacionados con la salud y el bienestar subjetivo, la valoración de los cuidados informales y la obtención de preferencias por estados de salud.

La tesis se estructura en cinco capítulos. El Capítulo 1 ofrece una introducción general, resumiendo algunos de los aspectos básicos de la evaluación económica de intervenciones sanitarias y exponiendo los antecedentes de cada uno de los temas tratados en los tres capítulos posteriores. Los Capítulos 2-4 presentan tres estudios empíricos independientes que conforman el núcleo de la tesis. En el Capítulo 2 se analiza la relación simultánea entre la calidad de vida relacionada con la salud (CVRS) y la satisfacción con la vida. El Capítulo 3 aborda el tema de la valoración monetaria de los cuidados informales, obteniéndose dicha valoración a partir de las preferencias de cuidadores informales y de no cuidadores. El Capítulo 4 examina el efecto del aprendizaje sobre la consistencia de las preferencias, empleando para ello tanto resultados en salud como resultados monetarios. Por último, en el Capítulo 5 se discuten los estudios presentados en los tres capítulos precedentes, destacando sus principales resultados, implicaciones, limitaciones y propuestas para futuras investigaciones. El contenido de los Capítulos 2-4 se resume a continuación.

Capítulo 2: La relación simultánea entre la CVRS y la satisfacción con la vida

El enfoque convencional adoptado en la literatura sobre la relación entre la salud autopercibida y el bienestar subjetivo se basa en la aplicación de modelos de regresión unidireccionales. En estos modelos generalmente se considera al bienestar subjetivo como dependiente de la salud autopercibida, sin tener en cuenta que, a su vez, la percepción del propio estado de salud se puede ver condicionada por el bienestar subjetivo (es decir, por cuán satisfecha y feliz se encuentre una persona). Frente a esta perspectiva unidireccional, el Capítulo 2 trata de esclarecer si la calidad CVRS[1] y la satisfacción con la vida (o componente cognitivo del bienestar subjetivo) se encuentran recíprocamente asociadas. Las variables de resultados –CVRS y satisfacción con la vida– se midieron empleando la tarifa de utilidades SF-6D (Brazier *et al.*, 2002) y la Escala de Satisfacción con la Vida (Diener *et al.*, 1985), respectivamente. La muestra objeto de estudio se compuso de 870 individuos, representativos de la población

[1] La CVRS es un concepto multidimensional que hace referencia a la autopercepción del estado de salud en sus vertientes física, psíquica y social.

general adulta española en términos de edad y sexo. Los resultados obtenidos ponen de manifiesto que, efectivamente, la relación entre la CVRS y la satisfacción con la vida es simultánea, de modo que, cuanto mejor percibe una persona su estado de salud, se encontrará más satisfecha con su propia vida y, a su vez, cuanto más satisfecha esté una persona, ésta tendrá una mejor percepción de su estado de salud. Asimismo, con independencia del enfoque adoptado (unidireccional o simultáneo), se observó que el efecto de la CVRS sobre la satisfacción con la vida es superior (casi un 20% tras tener en cuenta la simultaneidad) que el efecto opuesto.

Cabe destacar que la presencia de una relación simultánea entre la CVRS y la satisfacción con la vida implica que los modelos de regresión unidireccionales proporcionan estimaciones sesgadas e ineficientes del efecto de la CVRS sobre la satisfacción y del efecto contrario, pudiendo dar lugar, por tanto, a adoptar decisiones erróneas. Por este motivo, en el Capítulo 2 se cuantifica el sesgo cometido en caso de no tenerse en cuenta la interdependencia entre la CVRS y la satisfacción con la vida. En este estudio dicho sesgo se tradujo en una importante infravaloración del efecto de la satisfacción con la vida sobre la CVRS (en torno a un 18%) y, en menor medida, del efecto contrario (aproximadamente un 6%).

Es preciso matizar que los resultados comentados anteriormente se refieren a la totalidad de la muestra. El análisis comparativo entre hombres y mujeres y entre distintos grupos de edad mostró la existencia de heterogeneidad en términos de edad y sexo en la relación entre la CVRS y la satisfacción con la vida. En concreto, se halló evidencia de simultaneidad en el grupo de mujeres, individuos menores de 36 años y mayores de 49 años. En estos tres grupos el efecto de la satisfacción con la vida sobre la CVRS fue superior al efecto contrario. En cambio, en el grupo de hombres y de sujetos con edades comprendidas entre 36 y 49 años se encontró una relación unidireccional, siendo significativo el efecto de la CVRS sobre la satisfacción con la vida, pero no así el efecto opuesto.

La existencia de una influencia mutua entre la CVRS y la satisfacción con la vida (a nivel agregado) supone que las ganancias en CVRS no solo se pueden obtener por medio de intervenciones y políticas sanitarias, sino también gracias a otro tipo de acciones (por ejemplo, terapia cognitiva, risoterapia, arte terapia, actividades de voluntariado) y políticas públicas (tales como educación y empleo) que, a través de su impacto directo sobre la satisfacción, pueden inducir efectos colaterales positivos sobre la CVRS. A su vez, las intervenciones y políticas sanitarias también se configuran como potentes herramientas para la mejora del bienestar subjetivo, el cual es concebido como uno de los grandes objetivos de las políticas públicas. Por tanto, las políticas e intervenciones sanitarias y no sanitarias deberían estar estrechamente coordinadas con el fin de aprovechar las sinergias que puedan surgir entre ellas. Asimismo, el hecho de que la relación entre la CVRS y la satisfacción con la vida sea simultánea resalta la conveniencia de valorar de forma más amplia los resultados de las intervenciones sanitarias, no solo teniendo en cuenta aspectos estrictamente relacionados con la salud, sino también con la satisfacción con la vida.

Capítulo 3: La valoración monetaria de los cuidados informales a partir de las preferencias de cuidadores y no cuidadores

El Capítulo 3 presenta los resultados de un estudio de valoración contingente aplicado a la valoración monetaria de los cuidados informales. Dicha valoración se obtuvo a partir de las preferencias de dos muestras independientes: cuidadores informales (n = 202) y no cuidadores (n = 200), incluyéndose en este último grupo aquellos sujetos que ni son cuidadores informales ni personas en situación de dependencia. Cabe destacar que ningún estudio previo ha obtenido una valoración monetaria de los cuidados informales desde el punto de vista de no cuidadores, lo cual confiere un carácter inédito al Capítulo 3.

A todos los participantes en el estudio se les presentó un mismo escenario referido a una situación hipotética de cuidados informales y se les preguntó por la mínima compensación monetaria que exigirían –es decir, por su disposición a aceptar (DAA)– si tuvieran que cuidar durante una hora adicional al día a la persona dependiente descrita en dicha situación. En concreto, se formularon tres preguntas de DAA relativas al escenario hipotético: sin especificar la tarea a prestar en la hora adicional de cuidados informales (“$DAA_{general}$”); en relación a una hora extra realizando la tarea más molesta (“DAA_{peor}”); y en términos de una hora adicional dedicada a la tarea menos molesta (”DAA_{mejor}”). Mediante la comparación entre estas tres valoraciones, se analizó la consistencia lógica y la sensibilidad de las valoraciones en relación a las preferencias individuales entre distintas tareas de cuidados. Así, aquellos sujetos que indicaron el mismo valor en las tres preguntas fueron considerados “insensibles” o “invariantes”. Por otra parte, las respuestas reflejaron consistencia lógica cuando se cumplió la siguiente ordenación: $DAA_{mejor} < DAA_{general} < DAA_{peor}$. Por el contrario, fueron considerados inconsistentes aquellos encuestados que exhibieron al menos uno de los siguientes patrones de respuesta: $DAA_{mejor} > DAA_{general}$, $DAA_{mejor} > DAA_{peor}$ o $DAA_{general} > DAA_{peor}$.

En las dos muestras se obtuvo una pequeña proporción de ceros protesta (inferior a 8% en las tres preguntas de DAA)[2], y las valoraciones, a nivel agregado, cumplieron con el criterio de consistencia lógica previamente mencionado. Más concretamente, en la muestra de cuidadores, los valores medios/medianos de las DAA_{mejor}, $DAA_{general}$ y DAA_{peor} (en euros/hora) ascendieron a 5.3/4.5, 6.4/5.5 y 7.5/7, respectivamente; mientras que en la muestra de no cuidadores dichos valores (en euros/hora) se situaron en 5.6/5.5, 6.5/5.5 y 7.9/9, respectivamente. Las diferencias encontradas entre las distribuciones de las DAA obtenidas en las dos muestras no fueron significativas. A nivel individual, hubo una reducida proporción de inconsistencias (4,3% y 6,4% en la muestra de cuidadores y no cuidadores, respectivamente). No obstante, debido a la notable incidencia de respuestas invariantes (cerca de un 25% en cada grupo), algo menos de una tercera parte de cuidadores y no cuidadores fueron estrictamente consistentes.

Por otra parte, a los cuidadores informales también se les preguntó por la compensación monetaria que exigirían si tuvieran que cuidar a la persona dependiente a su cargo durante una hora adicional diaria, sin especificar la tarea a prestar ("DAA_{real}"). En este sentido, los valores medio/mediano de la DAA_{real} (5,2 €/4,5 €) fueron inferiores a los obtenidos en la pregunta de $DAA_{general}$ en la muestra de cuidadores (6.4 €/5,5 €), siendo significativa la diferencia entre las distribuciones de las dos variables. Aproximadamente la mitad de los cuidadores (52%) indicaron la misma cantidad en ambas preguntas, lo cual se relaciona con el hecho de que la mayoría de cuidadores (75%) reconocieron haber pensado en la persona a su cargo al responder a las preguntas referidas a la situación hipotética. Por tanto, estos resultados ponen de manifiesto que en gran medida las valoraciones de los cuidadores bajo el escenario hipotético se vieron influidas por su propia experiencia prestando cuidados informales. Si bien la DAA_{real} y la $DAA_{general}$ no son comparables (al referirse a dos situaciones de cuidados distintas), el hecho de que los valores medio y mediano obtenidos en la primera de las preguntas fueran inferiores a los obtenidos en la segunda sugiere que los aspectos positivos relacionados con los cuidados informales tuvieron un mayor peso al responder a la pregunta de DAA_{real} que cuando los cuidadores se enfrentaron al escenario hipotético.

En conclusión, los resultados de este estudio muestran que es factible obtener una valoración monetaria de los cuidados informales a partir de una muestra de no cuidadores (tal y como se puso de manifiesto en la reducida proporción de ceros protesta y de respuestas inconsistentes). El patrón de respuestas en este grupo fue muy similar al observado en la muestra de cuidadores, aunque las valoraciones de estos últimos parecen haber estado influidas por su propia experiencia como cuidadores informales.

[2] Los ceros protesta hacen referencia a las DAA nulas motivadas por cuestiones éticas. Dichas respuestas fueron excluidas del cómputo de los valores de las DAA.

Es importante tener en cuenta que los valores monetarios anteriormente mencionados no son generalizables porque se obtuvieron en referencia a una única situación hipotética. Por este motivo, sería interesante que en futuras investigaciones se obtuviera un conjunto de valores monetarios para una serie de situaciones relacionadas con los cuidados informales.

Capítulo 4: El efecto del aprendizaje sobre la consistencia de las preferencias

El fenómeno de la inversión de preferencias –o *preference reversal* (PR)– es un tipo de inconsistencia que se produce cuando dos métodos de obtención de preferencias que son teóricamente equivalentes (por ejemplo, valoración y elección) conducen a una ordenación diferente de las opciones evaluadas. En consecuencia, dicho fenómeno representa una quiebra del supuesto de invarianza, según el cual el método de elicitación de preferencias no debería influir en el ranking de las alternativas objeto de evaluación. El ejemplo clásico de PR comprende dos loterías, denominadas "P-bet" y "$-bet" (Lichtenstein and Slovic, 1971). La primera de ellas ofrece una alta probabilidad de conseguir un premio modesto, mientras que la segunda ofrece una menor probabilidad de obtener un premio de mayor cuantía. En un típico experimento de PR, los sujetos han de valorar cada lotería por separado y elegir entre la P-bet y la $-bet. Se pueden identificar dos tipos de inconsistencias: inversión de preferencias estándar y no estándar –*standard preference reveral* (SPR) y *non-standard preference reveral* (NSPR), respectivamente. SPR se produce cuando se elige la P-bet pero se asigna un mayor valor a la $-bet. NSPR es el patrón opuesto, es decir, se manifiesta cuando se elige la $-bet y se asigna un mayor valor a la P-bet.

Numerosos estudios han mostrado que PR es un patrón de comportamiento sistemático y frecuente, siendo SPR mucho más común que NSPR. Sin embargo, también existe cierta evidencia de que PR y otro tipo de anomalías tienden a atenuarse cuando los individuos cuando repiten varias veces las mismas tareas, reciben retroalimentación sobre las consecuencias de sus decisiones y, en algunos casos, incentivos. El estudio presentado en el Capítulo 4 se suma a la literatura sobre el efecto del aprendizaje sobre PR, siendo el primero en emplear tanto resultados de salud como resultados monetarios. En concreto, el principal objetivo de este estudio es examinar si la frecuencia de PR disminuye conforme los sujetos aprenden a través de la adquisición de experiencia con las tareas que realizan. Adicionalmente, en el Capítulo 4 también se examinan las diferencias entre hombres y mujeres en términos de PR, así como la influencia que el tipo de resultados empleados en las loterías (salud, dinero) tiene sobre PR.

El estudio se realizó en dos sesiones experimentales independientes, con el fin de distinguir el efecto de la mera repetición de tareas (primera sesión) del efecto combinado de la repetición de tareas junto con la retroalimentación sobre las consecuencias de las decisiones adoptadas (segunda sesión). En ambas sesiones se emplearon dos pares de loterías y se realizaron dos tipos de tareas: valoraciones –mediante la obtención del equivalente de certeza o *certainty equivalent* de cada lotería– y elecciones entre las dos loterías (P-bet y $-bet) de cada par. La muestra estuvo compuesta por 319 estudiantes de Grado de la Universidad de Murcia y fue dividida en tres grupos en función de la naturaleza de los resultados empleados: “dinero” (Grupo 1), “años de vida” (Grupo 2) y “días sin dolor de espalda” (Grupo 3).

En línea con previos experimentos sobre PR, los resultados obtenidos revelaron un marcado patrón asimétrico de PR, manifestándose SPR de forma mucho más frecuente que NSPR. En concreto, en término medio, un 50% y 53% de los encuestados incurrieron en SPR en las sesiones 1 y 2, respectivamente, mientras que la proporción de NSPR en las mismas sesiones se situó en 1,5% y 2,9%. La repetición de tareas en combinación con retroalimentación contribuyó a atenuar de forma significativa la presencia de SPR, al contrario que la repetición de tareas por sí sola. Como consecuencia de ello, se produjo un aumento significativo en el número de respuestas consistentes a lo largo de la segunda sesión, aunque dicha mejora se vio limitada debido a la tendencia creciente de NSPR en la misma sesión. De hecho, al final del experimento, la frecuencia de respuestas consistentes siguió siendo inferior a la de SPR, en término medio (38,6% frente a 48,1% de respuestas, respectivamente).

Por otra parte, en ambas sesiones las mujeres fueron más propensas que los hombres a incurrir en SPR. Además, se obtuvo una mayor proporción de SPR en los dos grupos que fueron expuestos a resultados en salud (especialmente en el grupo en el que los resultados de las loterías se expresaron en términos de años de vida) que en el grupo en el que se emplearon resultados monetarios. Este resultado destaca la necesidad de ser especialmente cuidadosos a la hora de diseñar estudios para la obtención de preferencias por estados de salud, con el fin de minimizar el riesgo de aparición de PR.

GENERAL ABSTRACT

The overall aim of this thesis is to examine new methodological approaches pertaining to the analysis of costs and outcomes of health care for the economic evaluation of health care interventions. In particular, these new approaches are applied in the context of the estimation of health and subjective well-being outcomes, the valuation of informal care and the elicitation of preferences over health states.

The thesis is structured into five chapters. Chapter 1 offers a general introduction, summarising some basic aspects of the economic evaluation of health care and providing the background of each issue addressed in the subsequent three chapters. Chapters 2-4 report three independent empirical studies which constitute the core of the thesis. Chapter 2 deals with the simultaneous relationship between health-related quality of life (HRQoL) and life satisfaction; Chapter 3 is concerned with the monetary valuation of informal care based on informal carers' and non-carers' preferences; and Chapter 4 examines the effect of learning on the consistency of preferences over health and monetary outcomes. Finally, Chapter 5 discusses the studies presented in the preceding three chapters, pointing out their main results, implications, limitations and suggestions for future research. The content of Chapters 2-4 is summarised below.

Chapter 2: The simultaneous relationship between HRQoL and life satisfaction

The standard approach adopted in the literature on the relationship between self-perceived health and subjective well-being relies on the application of unidirectional regression models. In general, these models consider that subjective well-being depends on self-perceived health, neglecting that, in turn, the self-perception of one's health can be conditioned by subjective well-being (that is to say, by how satisfied or happy a person is). Against this unidirectional perspective, Chapter 2 tries to clarify whether HRQoL[1] and life satisfaction (i.e. the cognitive component of subjective well-being are reciprocally associated. The outcome variables – HRQoL and life satisfaction– were assessed using the SF-6D index (Brazier *et al.*, 2002) and the Satisfaction With Life Scale (Diener *et al.*, 1985), respectively. The sample consisted of 870 individuals, being representative of the Spanish adult general population in terms of age and sex.

The results of the study revealed that the relationship between HRQoL and life satisfaction is actually simultaneous. This means that the better a person perceives his/her own

[1] HRQoL is a multidimensional concept which refers to the self-perception of one's health considering its physical, psychical and social domains.

health, the more satisfied he/she is, and, in turn, the more satisfied an individual is, the better he/she perceives his/her own health. Likewise, regardless of the approach adopted (unidirectional or simultaneous), the effect of HRQoL on life satisfaction was found to be stronger (by almost 20% after accounting for simultaneity) than the opposite effect.

The fact that HRQoL and satisfaction with life are simultaneously related implies that unidirectional regression models provide biased and inefficient estimates of the effect of HRQoL on life satisfaction and of the opposite effect, which can lead to make wrong decisions. For this reason, Chapter 2 quantifies the bias that occurs if the interdependency between HRQoL and life satisfactions is ignored. In this study, this bias gave rise to a substantial underestimation of the effect of life satisfaction on HRQoL (by around 18%) and, to a lesser extent, of the reverse effect (by approximately 6%).

It is important to note that the aforementioned results refer to the whole sample. The comparison between men and women and among three different age groups showed that the relationship between HRQoL and life satisfaction is heterogeneous in terms of age and gender. More specifically, there was evidence of simultaneity among women, respondents younger than 36 years and those older than 49 years. In these three groups, the effect of life satisfaction on HRQoL was stronger than the opposite effect. Conversely, in the male group and in the middle-aged (i.e. 36-49 years old) group, the relationship was found to be unidirectional, with the effect of HRQoL on life satisfaction being significant, unlike the opposite one.

The presence of a mutual influence between HRQoL and life satisfaction (at the aggregate level) entails that HRQoL gains can be achieved not only by means of health care interventions and policies, but also with other types of actions (for example, cognitive therapy, humour therapy, art therapy, volunteering) and policies (such as education and employment) which, through their direct impact on life satisfaction, can prompt positive side effects on HRQoL. In turn, health care interventions and policies can be effective tools for enhancing subjective well-being, which is conceived as one of the broad goals of public policies. This calls for a close coordination between health and non-health policies and interventions, with the aim of seizing synergies across them. Likewise, the simultaneity between HRQoL and life satisfaction stresses the appropriateness of assessing the outcomes of health care interventions more broadly, considering not only aspects strictly related to health, but also to life satisfaction.

Chapter 3: The monetary valuation of informal care based on carers' and non-carers' preferences

Chapter 3 presents the results of a contingent valuation study applied to the monetary valuation of informal care. This valuation was derived from the preferences of two independent samples: informal carers (n = 202) and non-carers (n = 200), where the latter group was composed of individuals who were neither informal carers nor care recipients. It should be noted that no prior study has obtained a monetary valuation of informal care from a non-carers' perspective, which adds an innovative character to Chapter 3.

All participants in the study were presented with a same hypothetical caring scenario and were asked to state the minimum amount of money they would require –that is, their willingness to accept (WTA)– if they had to spend one extra hour of informal care per day taking care of the dependant described in the hypothetical situation. Three different WTA questions were formulated in reference to that scenario: without specifying the task to be performed in the additional hour of informal care ('$WTA_{general}$'); in relation to one extra hour undertaking the most unpleasant task ('WTA_{worst}'); and in terms of one additional hour carrying out the least unpleasant task ('WTA_{best}'). By comparing these three WTA values, we analysed the logical consistency and the sensitivity of the valuations to the individual preferences over different caring tasks. Those respondents who expressed the same value in the three questions were regarded as 'insensitive' or 'invariant'. On the other hand, consistent respondents were those who met the following ranking: $WTA_{best} < WTA_{general} < WTA_{worst}$. Conversely, inconsistent individuals were those who exhibited at least one of the following patterns: $WTA_{worst} > WTA_{general}$, $WTA_{best} > WTA_{worst}$ or $WTA_{general} > WTA_{worst}$.

In both samples there was a small proportion of protest zeros (below 8% in the three hypothetical WTA questions)[2], and the valuations, at the aggregate level, satisfied the above-mentioned logical consistency criterion. To be exact, in the sample of carers, the mean/median WTA_{best}, $WTA_{general}$ and WTA_{worst} values (in euro/hour) amounted to 5.3/4.5, 6.4/5.5 and 7.5/7, respectively; whilst, in the sample of non-carers, the same values (in euro/hour) were 5.6/5.5, 6.5/5.5 and 7.9/9, respectively. Given that these values were similar in the two groups, no significant differences were found between the distributions of the WTA values obtained in the two samples. At the individual level, a limited proportion of carers and non-carers (4.3% and 6.4%, respectively) incurred in some inconsistency. Nevertheless, because of the considerable

[2] We considered that a respondent offered a protest zero when he/she rejected to receive any monetary compensation for one extra hour of informal care due to ethical objections. Protest observations were excluded from the study.

presence of invariant responses (nearly 25% in each group), a bit less than a third of carers and non-carers were strictly consistent.

On the other hand, informal carers were also asked how much money they would demand if they had to look after their own care recipient for one additional hour per day (without specifying the task to be undertaken) (‘WTA_{own}’). In this regard, the mean/median WTA_{own} values (€5.2/€4.5) were lower than the mean/median carers’ $WTA_{general}$ values (€6.4/€5.5), being significant the difference between the distributions of the two variables. Approximately a half of carers (52%) stated the same amount in both questions, which is related to the fact that most carers (75%) said they thought of their own care recipient when answering the three WTA questions referred to the hypothetical scenario. Therefore, these results suggest that carers’ hypothetical valuations were largely influenced by their own experience providing informal care. Albeit WTA_{own} and $WTA_{general}$ are not comparable (inasmuch they make reference to different caring situations), the fact that the values elicited in the former question were lower than those obtained in the latter one might suggest that the positive aspects of informal care played a greater role when carers faced their own situation than when they were presented with the hypothetical scenario.

In conclusion, the results of this study show that it is feasible to elicit a monetary valuation for informal care by asking a sample of non-carers (as it was evidenced by the low proportion of protest zeros and inconsistencies). The response profile in this group was very similar to that observed in the sample of carers, although the valuations in the latter group seemed to have been influenced by their own experience providing informal care.

It is worth noting that the monetary values mentioned above are not generalisable because they were obtained in reference to a single hypothetical situation. For that reason, it would be interesting to derive a set of monetary values for a range of caring situations in future studies.

Chapter 4: The effect of learning on the consistency of preferences

The preference reversal (PR) phenomenon is a type of inconsistency that occurs when two preference elicitation methods which, in theory, are equivalent (e.g. valuation and choice) lead to a different preference ordering of the alternatives being evaluated. In consequence, this phenomenon represents a violation of the procedural invariance assumption, according to which the preference elicitation method should not influence the ranking of the alternatives being evaluated. The classic example of PR entails two gambles, known as the ‘P-bet’ and the ‘$-bet’(Lichtenstein and Slovic, 1971). The former offers a high probability of a modest prize, whereas the latter offers a lower chance of a larger prize. In a typical PR experiment, respondents are asked to value the P-bet and the $-bet separately and to choose one of them. Two forms of inconsistencies can be identified: 'standard preference reversal' (SPR) and 'non-standard preference reversal' (NSPR). The former occurs when a subject chooses the P-bet but assigns a higher value to the $-bet. NSPR is the opposite pattern, that is, it happens when an individual chooses the $-bet but ascribes a higher value to the P-bet.

A large number of studies have found PR to be a systematic and frequent behavioural pattern, with SPR being far more common than NSPR. However, there is some evidence that PR and other types of preference anomalies tend to be attenuated as individuals have the opportunity to learn through the repetition of tasks, feedback on the consequences of their decisions and, in some cases, incentives. The study reported in Chapter 4 adds to the literature on the effect of learning on PR, being the first one to use health outcomes as well monetary outcomes. Specifically, the main objective of this study is to test whether the frequency of PR declines as subjects learn through acquisition of experience with the tasks they perform. Additionally, Chapter 4 also aims to examine gender differences in terms of PR, as well as the influence that the kind of lottery outcomes used (health and money) has on PR.

The study was conducted over two independent experimental sessions, in order to distinguish two learning effects: the effect of the mere repetition of tasks (first session) and the joint effect of repetition and feedback on the consequences of the decisions made (second session). Both sessions involved two pairs of gambles and two types of tasks: valuations –by eliciting the certainty equivalent values of each lottery– and binary choices between the two bets of each pair (P-bet and $-bet). The first session comprised three rounds of valuations and three rounds of choices for each pair of gambles. The second session was made up of three rounds of valuations for each lottery and a single choice for each pair. In this latter session, after each valuation, respondents observed 10 consecutive times the resolution of the risk associated with the gamble they had just valued (that is, they saw how that lottery could lead to either benefits or losses). The sample comprised 319 undergraduates at the University of Murcia (Spain) and was split into three groups, according to the nature of the outcomes used: 'money' (Group 1), 'years of life in good health' (Group 2) and 'days without back pain' (Group 3).

In line with previous experiments on PR, the results obtained revealed a pronounced asymmetric pattern, being SPR remarkably more frequent than NSPR. To be exact, on average, 50% and 53% of respondents incurred in SPR in sessions 1 and 2, respectively, whereas the rate of NSPR in the same sessions was only 1.5% and 2.9%. The repetition of tasks in combination with feedback contributed significantly to attenuate the frequency of SPR, unlike the repetition of tasks on its own. This led to a significant rise in the number of consistent responses over the second session, although this improvement was limited due to the upward trend in NSPR over the same session. Indeed, at the end of at the end of the experiment, the frequency of consistent responses was still significantly lower than that of SPR, on average (38.6% and 48.1% of responses, respectively).

On the other hand, in both sessions women were more likely than men to exhibit SPR. Furthermore, the rate of SPR was higher among those respondents who were exposed to outcomes (especially in the group where the outcomes of the gambles were expressed in terms of years of life) than among those who were presented with monetary outcomes. This result highlights the need to be especially careful when designing studies for the elicitation of preferences over health outcomes, with the aim of minimising the risk of appearance of PR.

CHAPTER 1

GENERAL INTRODUCTION

1.1. THE CONTEXT SURROUNDING THE TOPICS ADDRESSED IN THIS THESIS

1.1.1. The rationale for the economic evaluation of health care interventions

The fundamental problem of Economics is how to deal with scarcity, that is, how to satisfy needs (which are unlimited) with the existing resources (which are finite). In the health care sector scarcity entails serious consequences because it implies that the resources devoted to take care of a person are denied to other individuals and, consequently, there will always be people with unsatisfied health-related needs (e.g. patients who will go untreated). Therefore, it is an inexorable moral duty to seek the optimal allocation of the available health care resources. This is precisely the role of the economic evaluation of health care interventions[1], which is being increasingly used by health technology assessment agencies as a tool to ascertain which health care interventions should be prioritised and which not, in order to obtain the maximum health gains that can be achieved given the existing resource constraints (Canadian Agency for Drugs and Technologies in Health, 2006; Pharmaceutical Benefits Advisory Committe, 2008; National Institute for Health and Clinical Excellence, 2013).

The economic evaluation of health care interventions entails the comparison between two or more alternatives on the basis of their costs and consequences (Drummond *et al.*, 2005). In other words, an economic evaluation assesses whether or not the additional benefits generated by a health care intervention outweigh its opportunity costs (i.e. the benefits that could have been obtained from the next best alternative use of those resources) (Russell, 1992). This involves identifying, measuring, valuing and comparing the costs and outcomes of the different alternatives being examined (Drummond *et al.*, 2005). The problem is that the benefits of health care interventions (including health outcomes and broader effects, such as well-being gains) are not traded in the market and, therefore, no market prices exist for them. The same problem holds for many costs, such as time and productivity losses borne by patients and informal carers.

On this background, the general aim of this thesis is to explore new methodological approaches concerning the assessment of costs and outcomes for the economic evaluation of health care interventions. These new approaches are applied in the context of the estimation of health and well-being outcomes, the valuation of informal care and the elicitation of preferences

[1] The term ‘health care intervention’ includes pharmaceuticals, medical devices, diagnostic techniques, clinical and surgical procedures, health promotion activities, prevention programmes, as well as the organisational and support systems used in health care (Health Information and Quality Authority, 2010).

over health states. More specifically, Chapter 2 focuses on the simultaneous relationship between health-related quality of life (HRQoL) and life satisfaction; Chapter 3 deals with the monetary valuation of informal care on the basis of carers' and non-carers' preferences; and Chapter 4 tests the effect of learning on the consistency of preferences over health outcomes and money. The implications that the main findings obtained in these three chapters have for the economic evaluation of health care interventions are discussed in Chapter 5, which concludes the thesis. It is worth noting that the three studies reported in Chapters 2-4 represent a novelty in the field of health economics because each one adopts an approach that has not been previously tested. Hence, this work intends to contribute to the health economics literature and to broaden the debate around some of the methodological issues surrounding the assessment of the non-market items of an economic evaluation.

Since the core chapters (i.e. Chapters 2-4) tackle different issues related to some of the key components of the economic evaluation of health care interventions, the next subsection provides an overview on some basic aspects pertaining to them.

1.1.2. The basics of the economic evaluation of health care interventions

A central matter in an economic evaluation is the perspective under which it is conducted, insofar it determines the costs and outcomes that are deemed to be relevant for the analysis. The most coherent point of view with the maximisation of the expected health gains given the available resources is the societal perspective, which is the one of the society as a whole (Gold *et al.*, 1996; Drummond *et al.*, 2005). According to this view, all relevant costs and effects associated with an intervention should be taken into account, regardless of who bears the costs and who experiences the effects −the health care system, other sectors, patients, informal carers (i.e. members of the social environment of the patient, mainly relatives, who look after the patient), or the general public. The societal perspective is the most comprehensive one and it can be supplemented by other narrower perspectives, such as those of the patient, the health care provider (e.g. hospital), the payer (health care system or private insurer) or the public sector (which does not only include the health care sector, but also the rest of public bodies, such as social services and the educational system). Although the societal perspective is the optimal choice from the point of view of society's well-being, most national guidelines for the economic evaluation of health technologies still require the narrower health care system perspective (Claxton *et al.*, 2010). For instance, in the reference case, the National Institute for Clinical

Excellence in England (NICE) and Wales takes a limited perspective on costs –the only costs considered are those incurred by the National Health Service (NHS) and personal social services (PSS)–, whereas the perspective on outcomes encompasses all direct health effects for patients and, if relevant, for informal carers (National Institute for Health and Clinical Excellence, 2013).

In an economic evaluation, the concept of cost is understood as the opportunity cost that emerges when resources are consumed. The opportunity cost of a resource represents the forgone benefits from the next best alternative that is not chosen. That is, given the scarcity of resources, when a resource is allocated to a specific use, it is denied for other uses and this sacrifice causes an opportunity cost. The costs to be incorporated in an economic evaluation can be categorised into direct medical costs, direct non-medical costs and indirect costs (Luce *et al.*, 1996; Johnston *et al.*, 2001; Drummond *et al.*, 2005). Direct medical costs refer to the resources that are consumed within the health care sector as a direct consequence of the health care intervention (e.g. diagnosis tests, drugs, devices, medical care, in-patient treatment, hospital stay, etc.). Direct non-medical costs comprise the resources used outside the health care sector that are directly attributable to the intervention (e.g. costs borne by patients and informal carers, such as travel costs, time losses and out-of-pocket expenses). Finally, indirect costs include the consumption of resources that is not directly associated with the intervention (mainly productivity losses due to illness, disability or death). In order to be included in an economic evaluation, the costs items have to be valued in monetary terms, once they have been identified and measured in natural units. The economic theory postulates that, under certain circumstances, the opportunity cost of a resource can be reflected by its market price (Mishan and Quah, 2007). However, since many of the resources consumed as a result of a health care intervention are non-market 'goods', there is no price for them and, therefore, it is necessary to use some technique to impute a value to those resources.

On the other hand, the outcomes of a health care intervention (also referred to as effects, benefits or consequences) encompass both health-related and non-health related benefits. The latter category refers to productivity gains and improvements in quality of life others than health gains (e.g. due to changes in the way in which the treatment is administered). As mentioned earlier, if the societal perspective is adopted, all these effects should be computed as outcomes irrespective of who experiences them (Drummond *et al.*, 2005). In practice, however, economic evaluations are usually restricted to health-related outcomes for patients. The outcomes included in an economic evaluation can be measured in either monetary or non-monetary terms. Depending on the unit used to measure these outcomes, three main forms of economic evaluation are distinguished: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA), where the latter is a specific case of CEA.[2,3] In CBA the outcomes are monetised, whereas in the other two forms of economic evaluation the outcomes are expressed in non-monetary units. Below we briefly explain these three types of economic evaluation.

CBA is a type of economic evaluation, rooted in welfare economics, in which all benefits (as well as costs) are monetised. This allows a wide array of benefits to be computed, comprising both health-related and non-health related outcomes (Brazier *et al.*, 1999; Drummond *et al.*, 2005). Since both costs and benefits are expressed in the same unit, it can be directly determined whether a health care intervention is good value for money. This is the case when benefits are greater than costs, in other words, when the net benefit (i.e. benefits minus costs) is positive. Another advantage of CBA is that it is useful for comparing and setting priorities across different areas of public policy. Despite these advantages, the application of CBA in the field of health economics has been rather limited, mainly due to the difficulties in assigning a monetary value to health benefits and to the ethical objections against doing so (Cookson, 2003). Several methods can be used to monetise health care benefits, which can be categorised into revealed preference techniques and stated preference techniques. Revealed preference methods assume that preferences can be inferred by observing individuals' behaviour in markets closely related to the 'good' of interest. Some of these techniques are the travel cost method (Clarke, 1998; Puig-Junoy *et al.*, 1998) and the hedonic pricing method (Atkinson and Halvorsen, 1990). On the other hand, stated preference methods elicit respondents' preferences in reference to hypothetical markets or scenarios. These techniques include the contingent valuation (CV) method –which is mainly used by asking respondents to state their willingness to pay (WTP) for the good in question (Donaldson *et al.*, 2012)– and multi-attribute stated preference methods –which include both the conjoint analysis (CA) (Pinto *et al.*, 2000) and the discrete choice experiment (DCE) approaches (Ryan *et al.*, 2010).[4] More recently, the well-

[2] Although CUA is frequently referred to as CEA (e.g. Gold *et al.* 1996), in this thesis a distinction is made between both approaches (when appropriate) so as to avoid confusions.

[3] Other types of economic evaluation are cost-minimisation analysis, cost-consequence analysis and multi-criteria analysis.

[4] DCEs are frequently considered to be special cases of CA. However, there are actually considerably differences between these two forms of valuation techniques. On the one hand, in the CA respondents are aked to rank different alternatives or to rate them, whereas DCEs entail a number of choices between two or more options. On the other hand, the theoretical framework behind these methods is different. Whilst the CA is based on the conjoint measurement –which is not considered to be a sound behavioural paradigm of choice–, DCEs are based on random utility theory (Thurstone, 1927; McFadden, 1974)–

being valuation (WV) method has also been applied for the monetary valuation of health outcomes (Ferrer-i-Carbonell and Van Praag, 2002). Unlike the aforementioned techniques, this latter method is not based on preferences, but it relies on subjective well-being (SWB) data.

In CEA the outcomes are generally measured in natural (or physical) units, which can be final outcome measures (e.g. life years gained, number of cases prevented, number of deaths averted) or intermediate outcome measures (e.g. reductions in blood pressure, weight losses, improvements in bone mineral density). These measures are objective indicators of health that represent a benefit from a clinical perspective. However, there are also some CEAs where effectiveness is assessed on the basis of health-related quality of life (HRQoL) measures. HRQoL is a multidimensional construct that is closely related to the World Health Organization's definition of health (as representing 'a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity') (World Health Organization, 1948). It reflects how a person perceives his/her own health, considering various domains of health (e.g. physical, psychological and social functioning). Therefore, HRQoL goes beyond the traditional biomedical conceptualisation of health as absence of disease. The measures of HRQoL which are used in some CEAs are usually ordinal scales comprising different dimensions of health and, in turn, each dimension can include several items with different levels of severity, where the levels are measured on an ordinal scale. A distinction is often made between condition-specific and generic measures of HRQoL (Guyatt *et al.*, 1993). The former are used to assess the HRQoL of patients with a specific health condition. Some instruments included in this category are the Asthma Quality of Life Questionnaire (Juniper *et al.*, 1992), the Barthel Index (Mahoney and Barthel, 1965) and the Diabetes Quality of Life Questionnaire (The DCCT Research Group, 1988). On the contrary, generic measures of HRQoL are not restricted to a specific group of patients and, consequently, they can be used with different populations. Some of the most widely used generic measures of HRQoL are the Short Form 36 (SF-36) (Brazier *et al.*, 1992), the Short Form 12 (SF-12) (Ware *et al.*, 1996), the Sickness Impact Profile (Bergner *et al.*, 1976) and the Nottingham Health Profile (Hunt *et al.*, 1985).

It must be noted that, unlike the measures of HRQoL used in CUA (which will be described below), those used in CEA are not based on preferences. This means that the dimension scores of the above-mentioned instruments of HRQoL are calculated just by aggregating the different items that comprise each dimension, thus assigning the same weight to

---

which provides a comprehensive explanation of choice behaviour (Louviere *et al.*, 2010; Ryan *et al.*, 2010).

each item. Similarly, with some of these instruments an overall score is computed by adding the different dimensions scores, as if all dimensions were equally important (Brazier *et al.*, 1999). On the other hand, since the levels of the different items are represented on an ordinal scale (e.g. 1 = ‘very bad’, 2 = ‘bad’, 3 = ‘fair’, 4 = ‘good’, 5 = ‘very good’), for example, we can say that ‘bad’ is better than ‘very bad’, but we cannot quantify how much better is one option as compared with the other.

Another drawback of CEA is that, irrespective of the type of outcome measure used, this kind of economic evaluation is not useful for comparing interventions that differ in more than one outcome (e.g. increase in quality of life at the expense of a shorter length of life or vice versa) or interventions with different outcomes (e.g. reduction in blood pressure as opposed to reduction in blood glucose level; improvement in mobility *versus* improvement in mental health). In conclusion, CEA is of limited use to inform decisions about the efficient allocation of health care resources because it may lead to wrong (or suboptimal) decisions (Drummond *et al.*, 2005). For these reasons, CEA has found little acceptance among economists (Russell *et al.*, 1996; Brazier *et al.*, 1999), although it is extensively used in pharmacoeconomic studies. To overcome some of the limitations of CEA, a new variant of this type of economic evaluation, namely, the CUA approach, was developed. The peculiarity of CUA in comparison with the CEA is that in the former the benefits are measured in terms of quality-adjusted life years (QALYs). The QALY is a summary measure of health outcomes that combines into a single index changes in HRQoL and length of life (i.e. quality and quantity of life, respectively) (Kind *et al.*, 2009). A QALY represents the equivalent number of years that are lived in full health (Weinstein and Stason, 1977). Therefore, CUA is confined to HRQoL outcomes. For this reason, the QALY is not suitable to ascertain how much money should be spent on health care as compared with other public policies (Brazier *et al.*, 1999; Dolan, 2011). By contrast, given that the QALY is a standardised metric, its main advantage is that it enables comparisons across different diseases, programmes and populations, even across interventions that only have an impact on longevity and those that only have effects on quality of life (Weinstein *et al.*, 2009).

The number of QALYs is calculated by weighting the life years gained (or lost) as a result of a health care intervention by the patient’s HRQoL during that time. Leaving aside considerations of risk attitude and time preferences, the number of QALYs associated with any health state is calculated as follows (Pliskin *et al.*, 1980; Bleichrodt and Johannesson, 1997):

$$\text{Number of QALYs} = U(Q, T) = V(Q) \times T$$

where Q and T stand for the HRQoL of the health state in question and the number of years that it lasts, respectively, and $V(Q)$ is the value (or weight) attached to that HRQoL. This weight – which receives several names, such as quality (or HRQoL) weight, QALY weight, the '*Q*' part of the QALY or utility score– represents the utility attached to a given health state. This utility quantifies the strength of preferences (or desirability) for that health state (Drummond *et al.*, 2005). Therefore, QALYs are actually utility-adjusted life years (Richardson *et al.*, 1998), where utility is routinely assumed to be reflected by individuals' preferences for the health state in question. Consequently, in a CUA framework, those interventions that yield the most QALYs at the least cost should be given the highest priority (Weinstein, 1990).

The QALY weight is measured on a cardinal scale where 0 stands for 'death', 1 denotes 'full health', and negative values are assigned to those health states which are worse than death. To the extent that the QALY weight is measured on an interval scale, a change of a given magnitude always means the same, regardless of where it falls on the scale (Whitehead and Ali, 2010). In addition, it is assumed that a QALY has always the same value, irrespective of circumstances such as patients' age, socio-economic status or the severity of the health condition (Rodríguez and Pinto, 2000), although this is a controversial assumption, inasmuch it raises equity and fairness concerns (Tsuchiya, 2012). Limitations like this one have led some authors to question the 'welfaristic' grounds of the standard QALY approach, which focuses exclusively on individual preferences (Culyer, 1989; Hurley, 2000; Brouwer *et al.*, 2008). At present, the most elaborated 'extra-welfaristic' alternative to the conventional QALY foundations is that based on the Sen's 'capabilities' notion (Bleichrodt and Quiggin, 2013). This approach differentiates between 'functionings' and 'capabilities'. Functionings represent the outcomes that a person has actually achieved, whereas capabilities are the outcomes that a person is able to attain (Sen, 1993). In Sen's view, what really determines an individual's well-being is his/her capability set, rather than his/her functionings.

A pivotal issue in health economics revolves around the estimation of quality weights for health states (i.e. the '*Q*' of the QALYs), which entails the valuation of those health states. This valuation can be done directly or indirectly, being the latter option more frequent than the former. The most widely used methods for the direct valuation of health states are the standard gamble (SG) and the time trade-off (TTO). Both of them are commonly regarded as choice-based valuation techniques, in that respondents are asked to make trade-offs between quality of life and risk (in the SG) or between quality of life and longevity (in the TTO) (Dolan, 2000; Drummond *et al.*, 2005).

The SG elicits preferences under uncertainty and is based on the axioms of the von Neuman-Morgenstern expected utility theory (vNM EUT) (i.e. completeness, transitivity, independence and continuity) (von Neumann and Morgenstern, 1944). For chronic health states that are preferred to death, the SG requires respondents to choose between the certainty of living with a chronic health condition (*i*) for life (*t* years) and the uncertainty of a treatment with two possible outcomes: living in normal health for *t* years (with probability *p*) and dying immediately (with probability 1-*p*). Probability is varied until respondents are indifferent between the certain chronic health state and the risky treatment. At this point, the utility value for health state *i* is given by the probability *p* (Torrance, 1986).

Unlike the SG, the TTO obtains preferences under certainty (i.e. neglecting risk concerns). Although the TTO was developed as a pragmatic alternative to the SG and lacks theoretical foundation (Torrance *et al.*, 1972), the fact is that the TTO valuations can be regarded as analogous to a Hicksian measure of welfare change (Buckingham and Devlin, 2006). When the TTO is used for chronic health states that are regarded as better than death, respondents are asked to consider the length of time in full health (*x* years) that they consider to be equivalent to living for the rest of their lives (*t* years) with a specific chronic health condition (*i*). The indifference between the two alternatives is reached by varying *x*. At this point, the utility score for health state *i* is obtained by dividing *x* into *t* (Torrance, 1986).[5]

There is a third direct valuation method, the visual analogue scale (VAS), which involves rating each health state being assessed on a scale whose endpoints are 0 (denoting ‘death’) and 100 (representing ‘full health’). Nevertheless, the VAS is generally considered not to be an actual preference-based technique, in that it does not entail any choice (or trade-off) and, therefore, it is not generally recommended for use in economic evaluations (Stamuli, 2011). Instead, because of its simplicity, it is often used as a ‘warm up’ exercise before using other methods (Torrance *et al.*, 2001). There are, however, some researchers (Parkin and Devlin, 2006) who dispute the consensus against the VAS.

The indirect valuation of health states involves the use of multi-attribute health status classification systems, being the EuroQol 5D (EQ-5D) (EuroQol Group, 1990), the Short Form 6D (SF-6D) (Brazier *et al.*, 2002; Brazier and Roberts, 2004), the Health Utilities Index Mark 3 (HUI3) (Feeny *et al.*, 2002) and the Quality of Well-Being (QWB) Scale (Kaplan *et al.*, 1998) the most popular ones. All these instruments have a similar structure, comprising two parts: a

[5] A full description of how the SG and TTO operate when the chronic health state is worse than death can be found in Torrance (1986). Moreover, there is a recently developed variant of the classical TTO, which is able to assess both health states worse and better than death by using a uniform framing (Robinson and Spencer, 2006; Devlin *et al.*, 2011).

descriptive component and a valuation component. The former is a short questionnaire by means of which a health state is characterised in terms of several attributes or dimensions of health (e.g. mobility, mental health, self-care), where each dimension takes several levels of severity (e.g. 'a lot of problems', 'no problem'). In this way, any health state can be represented by the combination of the different attributes with their respective levels. The valuation component is an algorithm that attaches a utility score to each health state that can be described with the generic questionnaire. The value set for all those health states is also known as 'tariff' and has been previously obtained from a sample of the general public (trying to reflect the societal preferences), using some direct valuation method. For instance, the UK EQ-5D tariff (which is used by the NICE to determine the cost-effectiveness of different health technologies) was obtained from responses to TTO questions of a representative sample of over 3,000 members of the general public in the United Kingdom (Dolan *et al.*, 1996; Dolan, 1997). The same preference elicitation method was used to derive the Spanish EQ-5D tariff (Badia *et al.*, 2001). Conversely, the SF-6D tariff for the United Kingdom is based on the SG method (Brazier *et al.*, 2002; Brazier and Roberts, 2004). The Spanish tariff for the SF-6D was estimated using the probability lottery equivalent method, a variant of the SG that entails two risky alternatives (Abellán *et al.*, 2012).

Hitherto, we have used the concept of 'utility' as representing the strength of preferences (or desirability) for an outcome (Drummond *et al.*, 2005). This notion of utility has been called 'decision utility' (Kahneman *et al.*, 1997) and nowadays it is the prevailing view of utility among economists. On this view, what people prefer is good for themselves and, in consequence, it should be regarded as a benefit. Notwithstanding, it is worth pointing out that there is an alternative meaning of utility, which was indeed adopted for centuries. This is the notion of 'experienced utility', which refers to the hedonic experience of an outcome (i.e. what people like/dislike in the experience of their lives) (Dolan and Kahneman, 2008). According to this approach, the alternatives that people like the most will yield the greatest benefits (Edgeworth, 1881/1967).

Experienced utility is reflected by subjective well-being (SWB) measures. SWB is a broad construct which embraces three independent components: satisfaction with life, positive affect and negative affect (Diener, 2006). Life satisfaction is considered to be the cognitive component of SWB because it refers to the overall assessment of one's life, reflecting the gap between aspirations and achievements (Campbell *et al.*, 1976). The terms life satisfaction and happiness are frequently used interchangeably (Veenhoven, 1991), although happiness also has other meanings, such as 'a general mood, living a good life or the causes that make people happy' (Diener, 2006). Life satisfaction is frequently measured by asking respondents to rate on a certain scale how satisfied (or happy) they are with their life as a whole. Life satisfaction can also be assessed by asking individuals how satisfied they are with specific domains of their lives (health, job, family, financial situation, etc.), although it is preferable to use global assessments of life satisfaction because different subjects can ascribe a different weight to a same dimension of life (Diener *et al.*, 1985). Most economists and policy makers have focused on the cognitive component of SWB, which is in part explained because a number of national and international surveys (e.g. the British Household Panel Survey, the World Values Survey) now include life satisfaction questions. Overall, positive and negative affects constitute the affective (or emotional) component of SWB. Positive affect refers to pleasant moods and emotions (e.g. pleasure, euphoria, joy, interest, engagement, affection), whereas negative affect refers to unpleasant moods and emotions (e.g. anger, sadness, anxiety, worry, stress, frustration, guilt and shame, envy) (Diener, 2006). Affects can be measured by using survey questions –for instance, the Positive and Negative Affect Schedule (PANAS) (Watson *et al.*, 1988)– or more sophisticated techniques –such as the Day Reconstruction Method (DRM) (Kahneman *et al.*, 2004), where subjects provide moment-to-moment evaluations of how they feel while doing different activities through the day. An important advantage of SWB measures as compared with the QALY is that they can be used to assess the efficiency of different types of public interventions (e.g. education, environment, health care) and, therefore, to inform resource allocation decisions across them. By contrast, since the QALY only covers health-related aspects, this measure can be used only for the allocation of health care resources (Gandjour, 2001).

In conclusion, preferences reveal what people think they want, whilst experiences reflect what people really like or enjoy. If individuals were rational, they would prefer what they like the most and, therefore, decision utility and experienced utility would be equivalent. This assumption is implicit in many economic analyses, but in real life people sometimes want things that they do not like and vice versa (Dolan, 2008b; Dolan and Kahneman, 2008). In the case of CUA, there are some issues why preference-based measures of HRQoL and SWB measures may disagree. First, two of the most popular generic multi-attribute descriptive systems (EQ-5D and SF-6D) limit to health-related aspects, thus neglecting broader effects on SWB. Another widely used generic descriptive system, the HUI3, comprises a dimension of 'emotion', which refers to happiness. However, the other dimensions of the HUI3 are too focused on physical health. Overall, the EQ-5D and SF-6D attributes have a broader scope than the HUI3 dimensions. Comparing the EQ-5D and SF-6D descriptive systems, the former provides a narrower representation of HRQoL than the latter. For example, unlike the SF-6D, the EQ-5D does not include a dimension related to vitality. Thus, depending on the descriptive system used, important dimensions of HRQoL and SWB can be ignored.

Another important reason that explains the differences between preference-based measures of HRQoL and SWB measures refers to the way in which these measures are obtained. In general, preferences over health states are elicited by asking respondents of the general public TTO or SG questions. As previously explained, these techniques entails making choices between quality and quantity of life, given a hypothetical scenario. By asking the general public instead of patients, it is assumed that the societal preferences are captured. Conversely, SWB questions ask respondents about their own SWB and, therefore, no hypothetical choice is made.

Due to the aforementioned problems associated with the use of preference-based techniques for the valuation of health states, some researchers (Dolan, 2008a; Dolan and Kahneman, 2008) have firmly advocated that it would be more appropriate to value health states using more direct measures of the utility associated with those states (to be exact, SWB measures), directly reported by patients, as opposed to the elicitation of preferences (using either the SG or the TTO) from a sample of the general public. The main reasoning behind this claim is that preference-based measures do not properly reflect the impact of health states on people's lives, because circumstances often affect individuals quite differently from how they imagine them (Dolan and Metcalfe, 2012). This problem also exists when preferences are elicited by asking patients (instead of the general population), because in a preference elicitation question respondents often draw their attention to different things from those that would be really relevant when experiencing the health states being examined (Dolan, 2008a, 2011). Specifically, whereas the general public may overestimate the loss associated with a given health state (because they tend to emphasise its negative consequences), patients may underestimate that loss because of the adaptation to their health problems (Menzel *et al.*, 2002).

The preceding lines have outlined some basic concepts and ideas that will appear throughout this thesis. Next, sections 1.2-1.4 provide a more specific introduction to Chapters 2-4, respectively, by presenting the background and motivation for them, as well as the research questions that each of these chapters tries to answer.

1.2. THE RELATIONSHIP BETWEEN HEALTH-RELATED QUALITY OF LIFE AND LIFE SATISFACTION

As aforementioned, in CBA all outcomes derived from a health care intervention (including both health and non-health outcomes) can be incorporated in the analysis (depending on how the questions are posed). By contrast, although the measures of HRQoL used in CEA and CUA are broader than the traditional biomedical measures of health (e.g. biomarkers), their scope is still limited because they strictly focus on health outcomes. This fact may be problematic because the impact of many health care interventions goes beyond health and, therefore, the use of measures of HRQoL for the assessment of the outcomes of an intervention can neglect important consequences. This is particularly likely in the case of mental health interventions, public health strategies and elderly care (Drummond *et al.*, 2009).

Albeit the focus of economic evaluations (conducted in the form of CEA or CUA) is placed on health outcomes, health economists (as well as other researchers) show a growing interest in examining the effect of health on SWB. This is explained because health (independently of the measure used to assess it) is one of the most important determinants of SWB, as a large body of research has demonstrated. For instance, good self-perceived health has been found to be associated with higher levels of SWB (Gerdtham and Johannesson, 2001; Benyamini *et al.*, 2004; Helliwell and Putnam, 2004; Strine *et al.*, 2008; Watson *et al.*, 2010), whereas negative health outcomes (e.g. disabilities, diseases) lower SWB (Kendig *et al.*, 2000; Ferrer-i-Carbonell and Van Praag, 2002; Shields and Price, 2005; Wurm *et al.*, 2008; Graham *et al.*, 2011). Unsurprisingly, the positive association between SWB and health is stronger for mental health than for physical health (Dolan *et al.*, 2008). Likewise, SWB is more strongly correlated with measures of self-perceived health than with objective indicators of health (George and Landerman, 1984; Okun and George, 1984).

It is important to note that all the above-mentioned studies have treated SWB as dependent on health, which is logical because the former is a broader construct than the latter. Nonetheless, there is also wide evidence showing that health is conditioned by SWB. Indeed, it has been found that happier and more satisfied people have better perceptions of their own health (Al-Windi, 2005). The relationship from SWB to health is further supported by a number of longitudinal studies where happier and more satisfied people have been found to live longer (Lyubomirsky *et al.*, 2005; Veenhoven, 2008; Diener and Chan, 2011). In addition, other longitudinal studies have shown that dissatisfaction with life predicts both fatal unintentional injuries (e.g. traffic road accidents) (Kirkcaldy and Furnham, 2000) and fatal intentional injuries (e.g. suicide) (Koivumaa-Honkanen *et al.*, 2002), all-cause mortality (Maier and Smith, 1999), as well as work disability (Koivumaa-Honkanen *et al.*, 2004). Moreover, experimental studies have shown that induced positive mood (e.g. with humorous videos) boosts immune system (Lefcourt *et al.*, 1990), increases reaction to pain (Alden *et al.*, 2001), and restrains the increase in blood pressure when facing stressful events (Smith *et al.*, 2004).

In summary, regardless of the direction examined (from health to SWB or from SWB to health), the evidence shows that there is a strong positive correlation between health and SWB. This two-sided relationship suggests that health and SWB are simultaneously associated (i.e. that SWB and health cause each other). In order to shed light on this issue, Chapter 2 tests whether SWB and health are simultaneously related. In particular, given that the correlation between health and SWB is stronger when health is assessed with measures of self-perceived health than when objective indicators of health are used, we focus on the association between HRQoL and overall life satisfaction. As outcomes measures of HRQoL and life satisfaction, we use the preference-based SF-6D index derived from the Spanish tariff for the SF-6D (Abellán *et al.*, 2012) and the Satisfaction With Life Scale (SWLS) (Diener *et al.*, 1985), respectively. Our focus on the cognitive component of SWB is mainly explained because, in comparison with the affective component (positive and negative affects), life satisfaction is a less ambiguous (Campbell *et al.*, 1976) and more stable notion (Pavot and Diener, 1993).

It should be pointed out that the study of the simultaneous relationship between HRQoL and life satisfaction is not just a matter of curiosity. More important, the main reason why we investigate whether life satisfaction and HRQoL are simultaneously associated is that, if the relationship is actually simultaneous, the coefficients of the effect of life satisfaction on HRQoL and of the effect of HRQoL on life satisfaction will be biased if they are estimated using one of the unidirectional regression models conventionally employed in the literature on the relationship between health and SWB (such as ordinary least squares (OLS) and (ordered) probit/logit models).

The specific questions addressed in Chapter 2 are listed below:

- Is bidirectional the relationship between HRQoL and life satisfaction?
- If so, is there any bias if we model the relationship between them from a unidirectional perspective?
- Which effect is stronger: that from HRQoL to life satisfaction or the opposite one?
- Is the relationship between HRQoL and life satisfaction heterogeneous in terms of age and gender?

1.3. THE MONETARY VALUATION OF INFORMAL CARE

Informal care[6] is a complex and heterogeneous non-market 'commodity', in the sense that it involves the provision of different types of tasks and the amount of time devoted to these tasks can differ greatly depending, for instance, on the needs of the care recipient and on the availability of formal care (Van den Berg *et al.*, 2004). Furthermore, informal care entails both costs and effects. To a great extent, the costs of informal care result from the amount of time invested in caring. Since this time is sacrificed in other activities (such as labour market participation, family, social relationships or leisure), it gives rise to an opportunity cost (Hassink and Van den Berg, 2011). Additionally, informal carers often bear substantial financial costs because of outlays (e.g. in home adaptations to satisfy the care recipient's demands, drugs, medical devices, nappies, etc.), the inability to have a paid job or the reduction in hours of work. On the other hand, the provision of informal care may result in profound negative effects on carers' physical and mental health and SWB (Montgomery *et al.*, 1985; Schulz *et al.*, 1997; Coe and Van Houtven, 2009), albeit informal carers can also derive positive utility (e.g. fulfilment, satisfaction) from the fact of looking after a loved one (Brouwer *et al.*, 2005; Al-Janabi *et al.*, 2010)

Both costs and effects of informal care may vary over time. For example, a health care intervention targeted at the care recipient is likely to reduce the costs associated with informal care (since the care recipient will probably need less assistance) and to improve the carer's quality of life (due to the reduction in the time spent on informal care). As a result, ignoring the costs and effects of informal care in economic evaluations can lead to suboptimal resource allocation decisions (Van den Berg *et al.*, 2004). This holds both for economic evaluations of interventions targeted at patients and for those targeted at informal carers (e.g. respite care and support programmes) (Koopmanschap *et al.*, 2008).

Until now, only a few economic evaluations conducted have considered the costs and effects of informal care (Goodrich *et al.*, 2012), which may be related to the lack of consensus

[6] Informal care is the care provided, on a voluntary and ongoing basis, to people of one's social environment (family members, friends or neighbours) who need assistance in the activities of daily living (personal care, mobility, housework, etc.) due to health problems or aging. Although informal care is sometimes referred to as 'unpaid care' (in that informal carers do not receive a salary for the care provided), in some countries informal carers are entitled to receive some kind of financial reward (e.g. cash benefits, personal care budget, etc.), as long as the carer and/or the care recipient meet certain criteria. Therefore, in this thesis, the conceptualization of informal care is open to those carers who receive a compensation for the care they provide.

about how best to incorporate them in the analysis (Van den Berg *et al.*, 2004). Overall, the inclusion of informal care in economic evaluations can be made in two ways: a) by using a method that captures both costs and outcomes; or b) by separating costs and outcomes, valuing the former with a partial valuation method, and the latter in terms of carers' HRQoL or care-related quality of life (Goodrich *et al.*, 2012). The first option is compatible with all types of economic evaluations, while the latter is possible in CUA and multi-criteria analysis (Koopmanschap *et al.*, 2008).

Traditionally, the time devoted to provide informal care has been valued using either the opportunity cost method (Liu *et al.*, 2002; Van den Berg *et al.*, 2006) or the proxy good (or replacement) method (Oliva and Osuna, 2009; Sabes-Figuera *et al.*, 2010). Both of them are partial valuation methods because they consider only the cost of the time invested in caring, thus ignoring other types of costs as well as the positive and negative effects that emerge from the provision of informal care. Furthermore, neither of these two methods properly reflects carers' and care recipients' preferences (McDaid, 2001; Van den Berg *et al.*, 2005a). Given these limitations, alternative methods have been applied for the monetary valuation of informal care. These include the contingent valuation (CV) method (Van den Berg *et al.*, 2005a; De Meijer *et al.*, 2010), conjoint analysis (CA) (Van den Berg *et al.*, 2008) and discrete choice experiments (DCEs) (Mentzakis *et al.*, 2010), as well as the well-being valuation (WBV) method (Van den Berg and Ferrer-i-Carbonell, 2007). Whilst the latter technique is based on carers' SWB, the other ones are stated preference methods. In theory, all these methods can provide a complete valuation of informal care (i.e. they can capture all costs and effects resulting from the provision of informal care).[7]

The extensive use of the opportunity cost method and, to a lesser extent, of the proxy good method in those economic evaluations which have accounted for informal care −as compared with stated preference methods and the WBV method− may be related to the fact that research on the application of these alternative methods for the valuation of informal care is still scarce, though growing. In consequence, there is no consensus on the best way of using these methods in that context. One of the issues concerning the monetary valuation of informal care that has received very little attention is the standpoint under which that valuation is made (i.e. who is asked to value informal care). This is an important matter because it conditions the choice of the valuation method (and vice versa, the valuation method conditions the viewpoint of the valuation). For example, the WV method can be applied only with a sample of actual

[7] Whether stated preference methods provide a complete or a partial valuation depends on the questions asked and the tasks considered (Van den Berg *et al.*, 2004).

informal carers (insofar this method is based on a direct valuation of informal carers' SWB). In contrast, stated preference methods can be used from different perspectives: informal carers, care recipients, the general public as a whole or subjects who are neither carers nor care recipients ('non-carers', for short). The most straightforward way of valuing informal care when a stated preference method is used consists of asking actual informal carers, because they are the individuals who directly bear the costs and experience the effects associated with the provision of informal care. This could explain why all existing applications of the CV method for the valuation of informal care have focused on the carer's perspective (Van den Berg *et al.*, 2005b; Gustavsson *et al.*, 2010). Additionally, a few CV studies have obtained a monetary valuation for informal care from both a sample of informal carers and a sample of care recipients (Van den Berg *et al.*, 2005a; De Meijer *et al.*, 2010). Another alternative that has never been tested is to elicit preferences for informal care from the general public. This is consistent with the societal perspective, since members of the general public are potential, actual or former informal carers or care recipients and they bear part of the indirect costs of informal care. Furthermore, compared with informal carers and care recipients, non-carers could provide more objective (but less informed) valuations for informal care (Van den Berg *et al.*, 2004). This could lower the risk of strategic or self-interested responses and the incidence of protest zeros. Notwithstanding, the valuation of informal care from a non-carers' perspective poses some problems. For instance, non-carers may not have well-defined preferences over informal care and, as a result, their valuations may be less reliable than those elicited from informal carers.

Given that no published research has obtained a monetary value for informal care from a non-carer's perspective, the main objectives of Chapter 3 are to derive a monetary value for informal care from a sample of non-carers and to compare this valuation with that elicited from a sample of actual informal carers. The valuation method used was the CV technique –to be exact, in the form of willingness to accept (WTA).

More specifically, Chapter 3 tries to answer the following questions:

- Is it feasible to obtain a monetary value for informal care based on the stated preferences of non-carers?
- Does this valuation differ from that elicited from a sample of informal carers?
- Is this valuation consistent with (and sensitive to) the strength of preferences over different caring tasks?
- Are carers' values for informal care in reference to a hypothetical caring situation influenced by their own experience as informal carers?

1.4. THE EFFECT OF LEARNING ON THE CONSISTENCY OF PREFERENCES

The most widely used paradigm in economic evaluations is the von Neuman-Morgenstern expected utility theory (vNM EUT) (von Neumann and Morgenstern, 1944), which is the standard theory of individual decision making under uncertainty (Starmer, 2000). For instance, the SG method, which is commonly used for the direct valuation of health states, is rooted in the axioms of this theory, as previously mentioned. Nevertheless, there is ample evidence showing that EUT does not provide a valid characterisation of individual choice behaviour (Stalmeier and Bezembinder, 1999; Bleichrodt *et al.*, 2007; Abellán *et al.*, 2009).

One of the failures of EUT is the 'preference reversal' (PR) phenomenon. In particular, PR is considered to be a violation of procedural invariance (Tversky *et al.*, 1990). According to this assumption, the preference ordering over different alternatives should be independent of the method used to elicit those preferences (Tversky, 1996). That is to say, procedural invariance assumes that strategically equivalent methods (e.g. choice and valuation) should lead to the same ranking of the options being evaluated. In practice, however, failures of procedural invariance have been extensively documented in the literature (Lichtenstein and Slovic, 1971; Tversky *et al.*, 1988; Cox and Grether, 1996; Bateman *et al.*, 2007; Braga *et al.*, 2009; Loomes *et al.*, 2010).

PR has been mainly investigated in relation to monetary gambles. The classic example of PR −which was first reported by Lichtenstein and Slovic (1971)− involves two lotteries: the P-bet and the $-bet. The P-bet offers a high probability of winning a modest amount of money, whereas the $-bet offers a lower chance of winning a larger prize. Respondents are asked to place a monetary value on each gamble (typically elicited as a minimum selling price). The usual finding is that most respondents choose the P-bet but, at the same time, they assign a higher value to the $-bet. This pattern is called 'standard PR' (SPR). Conversely, the opposite kind of inconsistency (i.e. choosing the $-bet and valuing the P-bet more highly than the $-bet), which is known as 'non-standard PR' (NSPR), is observed occasionally. For instance, in the study reported by Lichtenstein and Slovic, around 83%, 51% and 56% of respondents incurred in SPR in Experiments I, II and III, respectively, whilst the rate of NSPR was substantially lower (approximately 6%, 27% and 11%, respectively). This asymmetric pattern cannot be explained by response error alone, suggesting that PR is a systematic preference anomaly rather than a random one (Cox and Grether, 1996; Braga and Starmer, 2005).

Although scarcer, there is also evidence of PR regarding health outcomes. Different forms of PR has been identified in the health economics literature, which can be categorised into: choice-judgment reversals (Sumner and Nease, 2001; Stalmeier and Verheijen, 2012; Oliver, 2013a, 2013b); judgment-judgment reversals (Oliver, 2003; Pinto and Abellán, 2005; Bleichrodt *et al.*, 2007); and choice-choice reversals (Oliver, 2006; Bleichrodt and Pinto, 2009). Chapter 4 of this thesis provides an overview of the literature on each of these types of PR. These examples cast serious doubts about the presumption that the preference ordering over different health outcomes and interventions is stable and independent of the procedure used to elicit those preferences (Oliver and Sorenson, 2008). Despite that, health economists use an array of preference-based methods –including choice-based methods (e.g. DCEs), matching procedures (e.g. SG, TTO), and monetary valuation techniques (e.g. WTP)– and implicitly assume that procedural invariance holds (Oliver, 2013a).

The substantial frequency of PR undermines the reliability of stated preference methods and raises the basic question of which technique best reflects the true underlying preferences (Braga and Starmer, 2005; Braga *et al.*, 2009). Furthermore, it challenges the pillars of standard economic theory, which rests on a number of assumptions about preferences, being procedural invariance one of them. What is more, in view of the constructed preference approach, PR is proof that preferences do not exist, but they are instead constructed depending on how the decision tasks are designed and framed (Slovic, 1995; Lichtenstein and Slovic, 2006). A different view is that of the discovered preference hypothesis (Plott, 1996). According to this approach, individuals own a set of stable underlying preferences, which are prior to the preference elicitation exercise. However, when subjects face unfamiliar tasks, their preferences may be blurry and their behaviour can be influenced by a number of biases. The discovered preference hypothesis argues that the stated preferences tend to converge with the underlying preferences as individuals discover what it is in their own interest to do, through a process of learning by means of repetition of tasks, feedback on the consequences of their decisions and incentives. This hypothesis is supported by some evidence showing that PR and other behavioural anomalies tend to subside in certain environments where learning is prompted (Cox and Grether, 1996; Shogren *et al.*, 2001; Loomes *et al.*, 2003; Braga *et al.*, 2009). In this regard, some researchers have stated that anomalies are economically significant only if they persist even after individuals have repeated the same tasks several times, have received feedback on the consequences of their decisions as well as incentives (Binmore, 1994, 1999).

Along these lines, Chapter 4 investigates the effect of learning on PR, using three types of outcomes: money, years of life and pain-free days. In all cases, two learning effects are distinguished: the effect of repetition alone and the combined effect of repetition and feedback.

In particular, this chapter tackles the following questions:

- Is the frequency of PR attenuated as individuals acquire experience with the tasks they perform (through repetition) and learn about the consequences of their own decisions (through feedback)?
- Does learning lessen the typical discrepancy between SPR and NSPR? If so, does the standard asymmetric pattern of PR evolve towards a non-standard asymmetric pattern (with NSPR being more frequent than SPR)?
- Is PR sensitive to the nature of the outcomes used (health, money)?
- Are there gender differences in terms of PR?

CHAPTER 2

ANALYSING THE SIMULTANEOUS RELATIONSHIP BETWEEN LIFE SATISFACTION AND HEALTH-RELATED QUALITY OF LIFE

Chapter based on:

Garrido, S., Méndez, I., & Abellán, J. M. (2013). Analysing the simultaneous relationship between life satisfaction and health-related quality of life. *Journal of Happiness Studies*, *14*(3), 1813-1838.

ABSTRACT

This chapter aims to examine whether life satisfaction and health-related quality of life (HRQoL) are simultaneously related, as well as to quantify the bias that occurs if simultaneity is not accounted for. The study sample consisted of 870 respondents, representative of the Spanish adult general population. Using a simultaneous equations system –with the Satisfaction With Life Scale (SWLS) and the SF-6D index as outcome variables–, we found a simultaneous association between life satisfaction and HRQoL, although this relationship is heterogeneous in individual characteristics such as age and sex. More important, the fact of estimating the relationship between life satisfaction and HRQoL under a unidirectional approach severely underestimates the effect of life satisfaction on HRQoL and, to a lesser degree, the reverse direction effect. In consequence, policy decisions intended to improve satisfaction with life or HRQoL can be wrong if they rely on unidirectional estimates. Another relevant implication of this research is that, as a result of the simultaneous relationship between life satisfaction and HRQoL, not only health interventions may increase satisfaction with life, but also policies that improve life satisfaction can lead to positive side effects on HRQoL.

2.1. INTRODUCTION

An increasing number of social scientists argue that the progress of a society should be assessed and monitored not only in terms of material well-being, but also by people's evaluations and feelings about their lives, that is, by considering subjective well-being (SWB) data (Diener and Seligman, 2004; Diener, 2006; Kahneman and Krueger, 2006; Dolan and White, 2007; Stiglitz *et al.*, 2009). Although it is still not clear how best to incorporate the concept of SWB into policy making, it is worth noting that some countries –such as the United Kingdom, France and Canada– are starting to seriously consider the idea of using SWB data to inform policy decisions and to evaluate public policies (Stiglitz *et al.*, 2009; Helliwell, 2011). Furthermore, SWB questions are now inserted into large national and international surveys.[1] In most cases, these questions are intended to measure the cognitive component of SWB, that is, life satisfaction[2] (Veenhoven, 1984; Diener *et al.*, 1985). However, SWB is a broader concept than the cognitive evaluation of one's life; specifically, it is made up of two additional components: positive and negative affects. Positive affect refers to the presence of pleasant moods and emotions (e.g. pleasure, affection, euphoria, joy, interest and engagement). Conversely, negative affect denotes the presence of unpleasant moods and emotions (e.g. anger, sadness, anxiety and worry, stress, frustration, guilt and shame, envy) (Diener, 2006). Positive and negative affects taken together constitute the affective or emotional component of SWB (Diener, 1984). The conceptualization of SWB as composed of three factors has received consistent empirical support (Lucas *et al.*, 1996; Arthaud-Day *et al.*, 2005).

In order to give a complete picture of SWB, it is necessary to consider its three components as well as the interdependencies and time-sequence among them (Kim-Prieto *et al.*, 2005). However, without disputing the relevance of the affective component of SWB, in this study we only focused on life satisfaction. There are several factors that support our choice. First, according to Campbell *et al.* (1976), life satisfaction is a less ambiguous concept than

[1] For instance, the World Values Survey (World Values Survey Association, 2005-2008) contains the following life satisfaction question: 'All things considered, how satisfied are you with your life as a whole these days? [1 dissatisfied … 10 satisfied]'. The Eurobarometer Survey (European Commision, 2011) asks a similar question, but using verbal labels instead of a rating scale.

[2] The term 'happiness' is frequently used as a synonym of life satisfaction. Indeed, Veenhoven proposed the same definition for both concepts ('happiness or life satisfaction is the degree to which an individual judges the overall quality of his life as a whole favourably') (Veenhoven, 1991). However, happiness also has other meanings, such as 'a general positive mood, living a good life or the causes that make people happy' (Diener, 2006). For that reason, as some researchers do, we avoided using the word 'happiness' in this article.

affect because it can be 'precisely defined as the perceived discrepancy between aspiration and achievement'. In addition, life satisfaction is a more stable notion than pleasant and unpleasant affects, which are frequently spontaneous reactions of short duration that arise as a result of a given situation (Pavot and Diener, 1993). Moreover, as Veenhoven (1993) pointed out, life satisfaction is the closest concept to the utilitarian notion of the 'good life' (or hedonic SWB), insofar it reflects 'the degree to which an individual judges the overall quality of his/her life as a whole favourably'. Many economists, psychologists and sociologists have also focused on this component of SWB (Diener *et al.*, 1985; Veenhoven, 1993; Di Tella *et al.*, 2001; Frey and Stutzer, 2002; Blanchflower and Oswald, 2004; Helliwell and Putnam, 2004).

Among other public policies (education, employment, environment, etc.), the measurement of SWB might be especially useful for health policy, since SWB and health are closely intertwined. Indeed, the World Health Organization defined health in terms of SWB, specifically as representing 'a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity' (World Health Organization, 1948). Thus, under this perspective, the measurement of health should go beyond objective indicators of health (e.g. biological measures, diagnosed diseases) because they have been found to be weakly correlated with SWB (George and Landerman, 1984; Okun and George, 1984). For this reason, subjective (or self-perceived) health measures are growingly used in social sciences –especially within the health economics literature (Böckerman and Ilmakunnas, 2009; García-Gómez, 2011; Cubí-Mollá and Herrero, 2012)– and in medicine (Grandy and Fox, 2012; Kalsekar *et al.*, 2012). A substantial body of empirical research has examined the association between life satisfaction and self-perceived health. Given that the former is a broader construct than the latter (Böckerman *et al.*, 2011), the prevailing empirical approach consists of considering life satisfaction as outcome variable and self-perceived health as independent variable. Studies adopting this perspective have revealed that those individuals with better (poorer) self-perceived health report higher (lower) levels of satisfaction with life (Gerdtham and Johannesson, 2001; Helliwell, 2003; Benyamini *et al.*, 2004; Bobinac *et al.*, 2010; Kwan, 2010; Böckerman *et al.*, 2011; Graham *et al.*, 2011). Although scarcer, other studies have focused on the opposite direction effect, that is to say, on the relationship from life satisfaction to self-perceived health (Al-Windi, 2005; Siahpush *et al.*, 2008). For instance, Siahpush *et al.* (2008) found that life satisfaction predicted self-perceived health and other health outcomes (absence of long-term conditions and physical health) after a two-year follow-up, even adjusting for health, socioeconomic and behavioural factors at baseline.

In summary, regardless of the direction examined (from self-perceived health to life satisfaction or vice versa), all the above-mentioned studies have found a positive association between life satisfaction and self-perceived health. In consequence, it might be hypothesised that the relationship between these two variables is simultaneous. Under this hypothesis, good self-perceived health would be beneficial for life satisfaction and, in turn, high levels of life satisfaction would improve the perception of one’s health. To the best of our knowledge, preceding research has not formally tested for simultaneity between life satisfaction (or SWB more generally) and self-perceived health. In econometrical terms, if the relationship between both variables is simultaneous, unidirectional models will provide biased and inconsistent estimates, underestimating or overestimating the true impact of life satisfaction on self-perceived health (and vice versa). Consequently, if unidirectional estimates were used to inform policy makers (for instance, in the context of the allocation of health care resources), they could lead to make wrong and suboptimal decisions.

Along these lines, the present study has two main aims. First, to ascertain whether life satisfaction and self-perceived health are simultaneously related. Since our findings show that the relationship is actually simultaneous, the second aim is to get an estimate of the simultaneity bias, that is to say, to quantify the extent to which the effect of life satisfaction on self-perceived health and the reverse effect are underestimated or overestimated if simultaneity is not accounted for. Additionally, the study also analyses whether the association between satisfaction with life and self-perceived health is heterogeneous in terms of individual characteristics such as age and sex.

It must be pointed out that in this study we assessed self-perceived health in terms of health-related quality of life (HRQoL), in order to take account of the multidimensional nature of health. To be exact, we used a preference-based (or utility-based) measure of HRQoL, namely the SF-6D index (Brazier *et al.*, 2002). Measures of this kind have a number of advantages as compared with non-preference based instruments of HRQoL, as will be explained in the section of Methods.

2.2. METHODS

2.2.1. Participants

The data used in this study were collected through a survey, which was conducted from March to May 2010 in the Region of Murcia, Spain. Participants were selected using a stratified random sampling design. Specifically, the target population (subjects aged 18 years and over who lived in private dwellings) was divided into strata following the age and sex structure of the

Spanish adult general public, according to the Living Conditions Survey 2009 (National Statistics Institute, 2009). A total of 1,020 potential respondents were initially approached by telephone through a random digit dialing process. Among them, 870 subjects agreed to take part in the survey and then they were interviewed face-to-face at their usual residence by trained interviewers.

2.2.2. The questionnaire

The questionnaire was divided into three sections, which collected information regarding life satisfaction, self-perceived health and personal characteristics. The full questionnaire is provided in Appendix 2A.

2.2.2.1. Life satisfaction

First, respondents' life satisfaction was assessed using the Satisfaction With Life Scale (SWLS). This instrument was developed by (Diener *et al.*, 1985) for the measurement of satisfaction with life as a whole. Therefore, the SWLS gives an overall summary score instead of a score for different domains of life satisfaction. The SWLS is a Likert scale with five items and seven levels of response in ascending order –from 1 ('completely disagree') to 7 ('completely agree'), with an indifference point ('neither agree nor disagree') set at level 4. The five items are worded as follows: 'In most ways my life is close to my ideal'; 'The conditions of my life are excellent'; 'I am satisfied with my life'; 'So far I have gotten the important things I want in life'; 'If I could live my life over, I would change almost nothing' (Diener *et al.*, 1985). The summary score of the SWLS is calculated by adding up the score for each item; therefore, it ranges from 5 to 35. According to the guideline for the interpretation of the SWLS scores (Diener *et al.*, 2006), they should be interpreted as follows:

- 31 – 35: Extremely satisfied
- 26 – 30: Satisfied
- 21 – 25: Slightly satisfied
- 20: Neutral (neither satisfied nor dissatisfied)
- 15 – 19: Slightly dissatisfied
- 10 – 14: Dissatisfied
- 5 – 9: Extremely dissatisfied

Although single-item life satisfaction questions are widely used within the field of SWB, they are prone to a number of potential psychometric problems (Diener, 1984). Conversely, a large number of studies conducted with different populations around the world have reported favourable psychometric properties for the SWLS (Diener *et al.*, 1985; Blais *et al.*, 1989; Pavot *et al.*, 1991; Lucas *et al.*, 1996; Arrindell *et al.*, 1999; Peterson *et al.*, 2005; Hultell and Petter Gustavsson, 2008). Particularly, these studies have found the SWLS to have strong internal consistency (or scale reliability), with Cronbach's alpha generally above 0.80 (Diener *et al.*, 1985; Peterson *et al.*, 2005). In the present study, the SWLS also showed high internal consistency (Cronbach's alpha: 0.84). However, it would have been slightly higher (Cronbach's alpha: 0.86) if the last item had been removed from the scale. On the contrary, items 2 and 3 made the major contribution to the reliability of the scale. Consistently, these items were the most strongly correlated with the overall scale (item-test correlation: 0.82 ad 0.84, respectively), whereas the lowest correlation was found for item 5 (item-test correlation: 0.75), as in (Diener *et al.*, 1985).

Test-retest reliability for the SWLS is also high for short temporal intervals but it goes down as the length of the period increases. For example, Pavot *et al.* (1991) found a one-month test-retest coefficient of 0.84, while Blais *et al.* (1989) reported a two-month test-retest coefficient of 0.64. The fact that the stability of the SWLS declines over time could be a weakness of the SWLS, although it might also reflect actual changes in respondents' life satisfaction as a result of relevant life events (Hultell and Petter Gustavsson, 2008). In addition, the SWLS has also shown adequate convergent validity with other related measures of life satisfaction, as well as good discriminant validity in relation to measures of emotional well-being (Diener *et al.*, 1985; Pavot *et al.*, 1991). Principal components analyses usually identify a single factor for the SWLS –which accounts for about 66% of the variance (Diener *et al.*, 1985; Arrindell *et al.*, 1991; Pavot *et al.*, 1991)–, suggesting that the SWLS taps a single construct.

2.2.2.2. Self-perceived health

Second, the questionnaire included the Short-Form 36 Health Survey (SF-36) (Ware and Sherbourne, 1992; Ware *et al.*, 1993). This instrument is a generic health profile (i.e. it can be used across different patient populations), being one of the most widely evaluated measures of HRQoL in the world (Garratt *et al.*, 2002). The SF-36 consists of 36 items, which are grouped into eight dimensions: 'physical functioning', 'role limitations due to physical problems', 'bodily pain', 'general health', 'vitality', 'social functioning', 'role limitations due to emotional problems' and 'mental health'. Responses to the SF-36 items are added up to obtain a score for each dimension. Two summary scores (physical and mental component summary scales) can also be derived (Ware *et al.*, 1994; Ware *et al.*, 2000).

Unfortunately, neither the scores for the SF-36 dimensions nor the summary scores are comparable and, in consequence, they cannot be combined into a single index. This limitation makes the SF-36 useless for economic evaluation purposes. What is more, since the SF-36 scores are not based on preferences, they cannot be used to compute quality-adjusted life years (QALYs) (Brazier *et al.*, 1999). Nevertheless, Brazier and colleagues (Brazier *et al.*, 1998; Brazier *et al.*, 2002) developed a way of obtaining a preference-based single index from the answers to the SF-36 items. These authors reduced the SF-36 into a six-dimensional classification health state system, called SF-6D, amenable to valuation (so that the general population could state reliable preferences on a subset of SF-6D health states), and estimated a scoring algorithm (based on those preferences) capable of predicting utility values for all possible SF-6D health states. Therefore, it is possible to derive preference scores from a SF-36 data set by mapping it onto the SF-6D, as long as the 11 items used in the SF-6D (see Table 2B.1 in Appendix B) have been answered by the respondents. In the present study we converted the SF-36 responses into the SF-6D index using the Spanish algorithm for the SF-6D.

The SF-6D shows good psychometric properties in terms of both reliability and validity (Coons *et al.*, 2000). For instance, a meta-analysis of studies that have used the Spanish version of the SF-36 (Vilagut *et al.*, 2005) reported that the reliability (Cronbach's alpha) of the SF-36 scales was above 0.70 in 96% of the evaluations, showing good discrimination among severity groups, moderate correlations with clinical indicators and high correlation with other instruments of HRQoL. The SF-6D descriptive system also showed good psychometric properties in our sample. Specifically, its internal consistency was quite high (Cronbach's alpha: 0.82). Moreover, the six dimensions of the SF-6D contributed to the reliability of the instrument, with item-test correlations ranging from 0.77 ('pain' dimension) to 0.82 ('mental health' dimension).

Albeit the psychometric properties of the SF-6D are similar to those of the EQ-5D (Walters and Brazier, 2003; Marra *et al.*, 2005; Petrou and Hockley, 2005), utility values generated by these two instruments of HRQoL are not interchangeable. On the one hand, the EQ-5D suffers from 'ceiling' effects, which are reflected by the presence of a large proportion of respondents at the top levels of the different dimensions of this instrument (Macran *et al.*, 2003; Bharmal and Thomas, 2006). Conversely, the SF-6D is affected by 'floor' effects, which are manifested by a high proportion of respondents at the bottom level of certain dimensions, particularly 'role limitations' (Longworth and Bryan, 2003; Brazier *et al.*, 2004; Ferreira *et al.*, 2008). As a result, for relatively severe conditions, the EQ-5D scoring algorithm provides higher utility scores than the SF-6D algorithm and, therefore, the latter gives rise to lower HRQoL gains (and, by extension, lower QALY gains) than the former. Notwithstanding, in the present study we used a SF-6D algorithm that lowers the minimum value (i.e. the 'floor') of the instrument (Abellán *et al.*, 2012)[3], presumably augmenting the agreement between the SF-6D and the EQ-5D. It is worth noting that this algorithm was estimated from a representative sample of the Spanish general population, whose characteristics were similar to those of the participants in this study.

2.2.2.3. Personal characteristics

Finally, the questionnaire concluded by asking respondents about some personal characteristics. We selected the most relevant factors from a review of the literature on the correlates of SWB (Dolan *et al.*, 2008) and HRQoL (Kind *et al.*, 1998; Franks *et al.*, 2003). These variables can be grouped into demographic characteristics (age and sex); socioeconomic factors (marital status, educational attainment, employment status, monthly income and place of residence); and other individual characteristics (such as personality traits and religious beliefs). With the exception of age (that was measured on a continuous scale), the remaining characteristics were measured using categorical questions. The response categories for each categorical variable will be detailed in the section of Results (see Table 2.1). Furthermore, although this study focused on self-perceived health, the questionnaire also included a multi-item question related to 'objective' health. In particular, respondents were asked if they had been diagnosed with some chronic disease by a physician from a list of the most prevalent ones in the Spanish adult general population. Albeit self-reported, we considered the answers to this question to be objective indicators of health because the question was worded in terms of those conditions that had been diagnosed by a health care professional.

To test whether our sample was representative of the Spanish general public, we compared both samples in terms of the aforementioned personal characteristics. To that end, we tested for statistical differences in means (for continuous variables) and proportions (for discrete variables) between the two samples by performing the two-tailed *t*-test and the chi-squared test, respectively.

[3] This algorithm was estimated using the probability lottery equivalent method –a variant of the standard gamble (SG) technique which involves two risky alternatives.

2.2.3. The model

Equations (1) and (2) represent the *a priori* bidirectional relationship between HRQoL and life satisfaction:

$$U_i = \alpha + \beta_1 S_i + X_i^{'} \delta + \varepsilon_{\mathrm{i}} \qquad (1)$$

$$S_i = \quad + \beta_2 U_i + W_i^{'} \quad + \quad {}_{\mathrm{i}} \qquad (2)$$

where U_i and S_i stand for the SF-6D index[4] and the SWLS score, respectively; α and γ are two intercepts; β_1 and β_2 are the coefficients of interest; X_i'and W_i' represent vectors of covariates with associated coefficients δ and τ, respectively; and ε_i and η_i are normally distributed random error terms.

Regarding covariates, we controlled for the above-mentioned set of personal characteristics. Age (the only continuous variable) was entered into the model in years and also in age squared divided by 100 (thereby allowing for a non-linear effect of age). The remaining regressors were entered into the model as categorical variables.

We first estimated equations (1) and (2) independently by means of ordinary least squares (OLS)[5] and then we controlled for simultaneity using a simultaneous equations approach. In particular, we consistently estimated the parameters of the system using the three-stage least squares estimator (3SLS) (Zellner and Theil, 1962). This estimator combines the conventional two-stage least squares (2SLS) with seemingly unrelated regressions. As a result, it provides consistent and more efficient estimates than the 2SLS estimator[6]. As previously indicated, OLS estimates will be biased and inconsistent if HRQoL and life satisfaction are jointly determined by unobserved individual characteristics.

---

[4] Those health states which are regarded as better than death are characterised by a utility index U_i above 0, with an upper bound at 1. Conversely, health states considered to be more undesirable than death take negative scores up to a lower bound of -1. Utility values of 0 and 1 are conventionally attached to death and full health, respectively (Torrance, 1986).

[5] Equation (2) was estimated by OLS since, in our sample, the SWLS summary score ranged from 6 to 35 and, thus, in practical terms, it is a continuous variable. In this regard, there is some evidence that it makes little difference whether one assumes cardinality or ordinality of SWB answers (Ferrer-i-Carbonell and Frijters, 2004).

[6] The 2SLS estimator regresses each endogenous variable on all the exogenous variables. Therefore, it uses the predicted values of these auxiliary regressions as instruments in the OLS regression of each equation in the system. The 3SLS estimator additionally takes into account the covariances across equation disturbances. Zellner and Theil (1962) provided a detailed description of the properties of the 3SLS estimator.

Given that the SF-6D index is bounded between -1 and 1 and the SWLS summary score is an integer falling within the interval 5-35, the estimates for equations (1) are not directly comparable with those for equation (2), because the range of variation of the two dependent variables differs to a great extent. To attain comparability among the estimates for both equations, we standardised the coefficients and standard errors in all regression analyses we performed. We made this standardisation by multiplying the original estimates by the standard deviation of the corresponding explanatory variable and dividing it by the mean of the dependent variable. The standardised coefficients stand for the effect, relative to the mean of the dependent variable, of a one standard deviation increase in the explanatory variable.

2.3. RESULTS

2.3.1. Background statistics

Since the 870 participants answered the full questionnaire, no one was excluded from the study. Table 2.1 displays the personal characteristics of the sample and compares them with those of the Spanish adult general population. Respondents were aged between 18 and 90 years (with mean age being close to 44 years) and there were roughly equal numbers of men and women. Almost two out of three respondents reported some chronic disease. Tests of differences in means/proportions between our sample and the Spanish general public are summarised in the last column of Table 2.1. At the 5% significant level, we found significant differences only in terms of educational attainment, which was lower in our sample.

Table 2.2 provides information pertaining to life satisfaction and self-perceived health, including the mean values of the SWLS summary score and the SF-6D index. Additionally Figures 2.1 and 2.2 display the distributions of these two variables. Regarding life satisfaction, the SWLS summary score ranged from 6 to 35, with a mean value of 25.4. Following the aforementioned criteria for the interpretation of the SWLS scores (Diener *et al.*, 2006), most subjects fell in the range of ‘satisfied’ (37.9%) or ‘slightly satisfied’ (25.2%). There were also a considerable proportion of ‘extremely satisfied’ respondents (17.9%). By contrast, the percentage of respondents with scores at the bottom of the scale was much lower: whilst 10.2% of the sample lied within the range of ‘slightly dissatisfied’, only 4.0% and 0.8% of individuals were ‘dissatisfied’ and ‘extremely dissatisfied’, respectively. Finally, 3.9% of respondents were ‘neither satisfied nor dissatisfied’. As far as self-perceived health is concerned, the SF-6D index ranged from -0.13 to 1, with a mean value of 0.767. The fact of having obtained negative utility scores denotes that some subjects –in particular, only five respondents or 0.6% of the sample– considered their health state to be worse than death. Most participants rated their health as very good (31%) or good (37%), while 13.6% perceived their health to be excellent and 16.4% to be fair. Only 2% of respondents described their health as poor.[7]

Table 2.1. Sample and Spanish population personal characteristics

	Sample (N = 870)	Spanish population[a]
Mean (SD) age	43.8	46.1
Female (%)	50.1	50.9
Marital status (%)		
Single	33.7	31.8
Married/living together	56.8	56.7
Divorced	4.4	4.1
Widowed	5.2	7.4
Educational attainment (%)		
Illiterate or low	47.6	30.1*
Medium	34.4	45.1*
High	18.1	24.7*
Employment status (%)		
Employed/self-employed	48.4	43.7
Unemployed	12.1	16.5
Homemaker	14.3	25.8
Student	10.2	
Retired	15.1	14.0
Monthly income (%)		
Less than €900	26.9	23.5
€901-1,500	26.3	28.0
€1,501-2,000	22.4	19.4
€2,001-3,000	17.4	19.7
More than €3,000	6.9	9.4
Place of residence (%)		
Urban area	27.4	N/A
Rural area	72.6	N/A
Personality (%)		
Optimist	37.2	N/A
Realist	54.5	N/A
Pessimist	8.3	N/A

[7] These proportions were derived from the responses to the first question of the SF-36 instrument: 'In general, would you say your health is: excellent/very good/good/fair/poor?'

Table 2.1 (continued)

	Sample (N = 870)	Spanish population[a]
Religion (%)		
Non-believer	18.1	N/A
Non-praticing believer	53.3	N/A
Practicing believer	28.6	N/A
Chronic diseases (%)		
Allergy	14.2	11.6
Arthritis/rheumatism	17.4	21.7
Bronchitis	5.7	4.0
Diabetes	7.0	6.0
Migraine	18.2	14.5
Prostate	3.6	2.9
Musculoskeletal	43.6	N/A
Poor circulation	16.8	N/A
Cardiovascular	5.5	N/A
Mental disorders[b]	12.3	13.8
Hypertension	11.8	18.9
Cholesterol	13.3	14.2
None	35.2	N/A

[*] Indicates that the difference between the sample and population averages is statistically significant at the 5% significance level. N/A: Not available. SD: Standard deviation.
[a] The statistics for the Spanish population were calculated using the National Health Survey 2006 and the Living Conditions Survey 2009 (National Statistics Institute, 2006, 2009, respectively).
[b] Mental disorders: Nervous disorders and depression.

Table 2.2. Life satisfaction and self-perceived health statistics

Life satisfaction[a]	N = 870	*Self-perceived health*	N = 870
Extremely satisfied	17.9	Excellent	13.6
Satisfied	37.9	Very good	31.0
Slightly satisfied	25.2	Good	37.0
Neither satisfied nor dissatisfied	3.9	Fair	16.4
Slightly dissatisfied	10.2	Poor	2.0
Dissatisfied	4.0	Mean (SD) SF-6D index	0.8 (0.2)
Extremely dissatisfied	0.8		
Mean (SD) SWLS summary score	25.4 (5.6)		

Data for the variables listed in Table 2.2 are not available at the Spanish population level.
[a] We followed Diener *et al.* (2006) for the interpretation of the SWLS summary scores.

Figure 2.1. Distribution of the SF-6D index

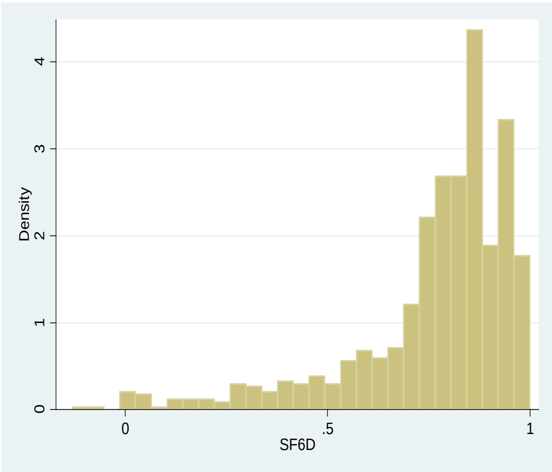
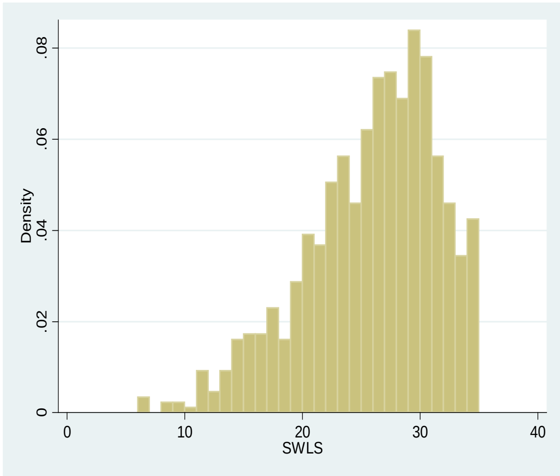


Figure 2.2. Distribution of the SWLS summary score



2.3.2. The unidirectional approach

The first three columns of Table 2.3 present the OLS standardised estimates for equation (1) (i.e. the SF-6D equation). We considered three different models. First, in column (1) we included the SWLS summary score as an explanatory variable. Second, in column (2) we split this scale into its five items. Finally, in column (3) we removed the non-significant items of the scale. The standardised estimates for equation (2) (i.e. the life satisfaction equation) are provided in column (4).

As shown is Table 2.1, our sample deviated significantly from the Spanish adult general population norms in terms of educational attainment. In order to ensure the population validity of our estimates, we used corrective weights to change the distribution of educational levels in the estimation sample for that in the Spanish general public.

2.3.2.1. The relationship from life satisfaction to HRQoL

The estimates in column (1) of Table 2.3 show that respondents who were more satisfied with their lives reported better HRQoL. In particular, the standardised coefficient of the SWLS summary score indicates that, all else equal, a one point increase in the standard deviation of the SWLS summary score results in a 0.054 point increase in the mean of the SF-6D index. This is equivalent to an elasticity of 0.253 –i.e. a one percent increase in life satisfaction is associated with a 0.253 percent increase in the SF-6D index.

Table 2.3. Unidirectional approach. OLS standardized estimates[a,b]

	SF-6D index						SWLS summary score	
	(1)		(2)		(3)		(4)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant	0.485***	0.067	0.462***	0.067	0.462***	0.065	25.135***	2.298
SWLS summary score	0.054***	0.007						
SWLS1			0.003	0.009				
SWLS2			0.056***	0.009	0.055***	0.009		
SWLS3			0.021**	0.009	0.018**	0.008		
SWLS4			0.000	0.008				
SWLS5			-0.012	0.007				
SF-6D index							0.074***	0.009
Age	0.196***	0.044	0.186***	0.050	0.190***	0.050	-0.176***	0.055
$Age^2/100$	-0.199***	0.043	-0.186***	0.050	-0.192***	0.050	0.191***	0.055
Female	-0.008	0.007	-0.007	0.007	-0.007	0.007	0.012	0.008
Single	0.024***	0.009	0.026***	0.009	0.025***	0.009	-0.035***	0.010
Divorced	0.002	0.006	0.001	0.006	0.002	0.006	-0.030***	0.007
Widowed	-0.016**	0.007	-0.012*	0.007	-0.013*	0.007	-0.012	0.008
Illiterate	-0.014	0.010	-0.013	0.009	-0.012	0.009	-0.037***	0.010
Low	-0.003	0.009	-0.003	0.009	-0.002	0.009	-0.029***	0.010
Middle	-0.001	0.009	-0.001	0.008	-0.001	0.008	-0.029***	0.009
Unemployed	0.004	0.007	0.003	0.007	0.003	0.007	-0.011	0.007
Homemaker	0.015*	0.008	0.015*	0.008	0.015*	0.008	-0.008	0.009
Student	0.013	0.008	0.009	0.008	0.010	0.008	0.003	0.009
Retired	-0.002	0.011	-0.004	0.011	-0.003	0.011	-0.004	0.012
Less than 900 €	-0.008	0.010	-0.007	0.010	-0.007	0.010	-0.028***	0.010
901-1,500 €	0.010	0.009	0.011	0.008	0.012	0.008	-0.038***	0.009
1,501-2,000 €	-0.002	0.008	-0.003	0.007	-0.003	0.007	-0.003	0.008
Rural area	0.003	0.007	0.005	0.007	0.005	0.007	0.003	0.007
Realistic	-0.006	0.006	-0.006	0.006	-0.007	0.006	0.008	0.007
Pessimistic	-0.009	0.007	-0.008	0.007	-0.008	0.007	-0.020***	0.007
Non-practicing believer	-0.006	0.008	-0.007	0.008	-0.007	0.008	0.031***	0.009
Practicing believer	-0.013	0.009	-0.014	0.009	-0.015*	0.009	0.036***	0.010
Allergy	-0.013**	0.006	-0.012*	0.006	-0.011*	0.014	-0.006	0.007
Arthritis/rheumatism	-0.030***	0.008	-0.032***	0.008	-0.031***	0.016	0.013	0.009
Bronchitis	-0.019***	0.007	-0.017***	0.006	-0.018***	0.021	-0.003	0.007
Diabetes	-0.003	0.007	-0.002	0.007	-0.002	0.020	-0.009	0.007
Migraine	-0.013*	0.007	-0.011*	0.007	-0.011*	0.013	0.001	0.007
Prostate	-0.003	0.007	-0.003	0.007	-0.004	0.027	0.010	0.007
Musculoskeletal	-0.039***	0.007	-0.040***	0.007	-0.039***	0.011	-0.005	0.008
Poor circulation	-0.017**	0.007	-0.016**	0.007	-0.017**	0.015	0.016**	0.008
Cardiovascular	-0.018***	0.007	-0.015**	0.007	-0.016**	0.023	-0.005	0.007
Mental disorders	-0.065***	0.007	-0.065***	0.007	-0.064***	0.017	-0.029***	0.008

Table 2.3 (continued)

	SF-6D index						SWLS summary score	
	(1)		(2)		(3)		(4)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Hypertension	-0.017**	0.008	-0.016**	0.007	-0.015**	0.017	0.000	0.008
Cholesterol	-0.004	0.007	-0.004	0.007	-0.004	0.015	0.005	0.007
N	870		870		870		870	
Adjusted R^2	0.539		0.557		0.558		0.271	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SWLS1-SWLS5: Items 1-5 of the SWLS.
Reference categories for discrete variables are: male; married/cohabiting; high educational level; employed/self-employed; more than €2,000 per month; optimistic; non-believer; with no chronic disease.
[a] The standardised coefficients and standard errors (SEs) were obtained by multiplying the original estimates by the standard deviation of the corresponding explanatory variable and dividing by the mean of the dependent variable.
[b] All regressions were estimated using corrective weights to change the distribution of educational attainment in the estimation sample for that in the Spanish adult general population.

Turning to the control variables, we found an inverted *U*-shaped relationship between age and the SF-6D index, with the latter increasing up to the age of 50, where a maximum was reached, and then falling down. Regarding marital status, compared with those married or living together, on average, single respondents scored a higher mean SF-6D index, while widowed respondents scored lower. Employment status, personality traits and religious beliefs were not significantly related to the SF-6D index. On the other hand, with the exceptions of cholesterol, diabetes and prostate problems, the remaining chronic diseases included in the model were negatively associated with the SF-6D index. The magnitude of the association was especially large for mental disorders (-0.065) and, to a lesser degree, for musculoskeletal problems (-0.039) and arthritis/rheumatism (-0.030). From columns (2) and (3), it must be seen that only the second and the third items of the SWLS –worded as ‘The conditions of my life are excellent’ and ‘I am satisfied with my life’, respectively– were found to be significantly associated with the SF-6D index. The standardised coefficient of the second item of the SWLS (0.056) was almost the same as that of the SWLS summary score (0.054), but it was three times larger than the standardised coefficient of the third item (0.021).

2.3.2.2. The relationship from HRQoL to life satisfaction

The estimates in column (4) of Table 2.3 attest that respondents who reported better HRQoL were more satisfied with their lives. More specifically, the standardised coefficient of the SF-6D index shows that, all else equal, a one point increase in the standard deviation of the SF-6D index leads to a 0.074 point increase in the mean value of the SWLS summary score, which is equivalent to an elasticity of 0.279.

The comparison between the standardised coefficients of the SWLS summary score and the SF-6D index in columns (1) and (4) of Table 2.3, respectively, reveals that the effect of life satisfaction on the SF-6D index (0.054) is 27% lower than the opposite direction effect (0.074). However, it must be pointed out that these estimates are biased if the life satisfaction and the SF-6D index are jointly determined by unobserved factors.

On the other hand, we found life satisfaction to be associated with all the demographic and socioeconomic factors used as control variables except with gender, employment status and place of residence. Our results stand for a *U*-shaped relationship between age and life satisfaction, indicating that younger and older respondents reported higher average levels of satisfaction with life, whereas middle-aged respondents reported the lowest average levels. In particular, the model predicts that the minimum value of life satisfaction is reached at the age of 45 years. It should be noted that the relationship between age and life satisfaction is just the opposite of that we found between age and the SF-6D index. Married respondents reported higher average scores of life satisfaction than single and divorced ones. Interestingly, we found average levels of life satisfaction to increase with the level of education, with those holding a university degree being, on average, more satisfied than those with any lower educational level. Likewise, we also found a non-monotonic and positive association between disposable income and life satisfaction. Regarding personality traits, pessimistic respondents reported lower average levels of life satisfaction than optimistic ones. Conversely, religious respondents (especially those who declared to be practicing believers) were, on average, more satisfied than non-believers.

Finally, we barely found evidence of a significant relationship between life satisfaction and chronic health diseases. The only exceptions were mental disorders and poor circulation. Notwithstanding, whilst the estimate for mental disorders was negative, that for poor circulation was positive, contrarily to what might have been expected.

2.3.3. The simultaneous equations model

The positive and highly significant correlation observed between the SWLS summary score and the SF-6D index under the two unidirectional approaches previously examined suggests that the relationship between life satisfaction and HRQoL can be bidirectional. To evaluate whether a simultaneous model of the two variables is called for, we tested for independence of equations (1) and (2) by estimating bivariate probit models whose dependent variables were discrete outcome measures of the SF-6D index and the SWLS summary score, respectively. To be exact, we created three dummies for each outcome variable, indicating whether the observed value lies below percentile 25th, below the median (or percentile 50th) or above percentile 75th. Then, we tested for pairwise independence between the two equations error terms using bivariate probit models.[8] The bivariate probit estimates are summarised in Table 2.4. Remarkably, the correlation between the error terms of the SF-6D index and the SWLS standardised measures was negative and significant at the 5% significance level in all the comparisons performed except for three, in which it was significant at the 10% significance level. This means that the unobserved factors that simultaneously affect HRQoL and life satisfaction exert an opposite direction effect on both variables. In consequence, we can conclude that the simultaneous model has empirical support.

Table 2.4. Bivariate probit estimates of the correlation between the error terms of the two equations

SF-6D indicators	SWLS summary score indicators		
	Lower than p25	Lower than median	Above p75
Lower than p25	-0.662	-0.868	-0.317
LR test *p*-value[a]	0.000	0.000	0.075
Lower than median	-0.828	-0.761	-0.327
LR test *p*-value[a]	0.000	0.000	0.067
Above p75	-0.966	-0.757	-0.227
LR test *p*-value[a]	0.000	0.000	0.073

The list of regressors is the same as in Table 2.3. p25 and p75 stand for the 25th and the 75th percentiles, respectively, of each variable.

[a] The likelihood ratio (LR) tests the null hypothesis of no correlation between error terms. It follows a chi-squared distribution with one degree of freedom.

[8] Martínez-Granado and Ruiz-Castillo (2002) suggested using this approach to examine the convenience of using a simultaneous equations model.

A simultaneous equations model is not identified unless there is at least one control variable that is statistically related to the outcome variable of one of the equations, but not to the outcome variables of the remaining equations in the system. The variables that satisfy the conditions for identification are known as exclusion restrictions. We identified the system by assuming that: (1) religion has a direct effect on satisfaction with life, but not on HRQoL; and (2) most chronic diseases only affect life satisfaction through their direct effect on HRQoL. These assumptions are grounded on the unidirectional estimates shown in Table 2.3 and, more importantly, on empirical evidence.

The first assumption is supported by evidence from research showing that people who become more religious over time experience long-term gains in life satisfaction, while those who become less religious suffer long-term losses (Heady *et al.*, 2008). Similarly, there is also evidence that life satisfaction is enhanced among those individuals who practice a religion that promotes social capital –such as the Catholic religion, the predominant one in Spain (Okulicz-Kozaryn, 2009). The aforementioned studies include numerous references that support the hypothesis that religion exerts a direct positive influence on life satisfaction. Conversely, the direct effect of religion on adult health is not so well documented. To a large extent, the existing evidence on the effect of religion on health behaviours and lifestyles refers to minority or conservative religious denominations that clearly differ from the Catholic religion.[9] For instance, Ferraro and Albrecht-Jensen (1991) found that religion may have both positive and negative effects on self-perceived health. Specifically, using a national sample of American adults, they found that, irrespective of age, those with a more conservative religious affiliation reported poorer health than those with a more liberal affiliation. Nevertheless, higher levels of religious practice were positively related to better health. Overall, in this study the positive effects of religion were stronger than the negative ones.

On the basis of these the above-mentioned studies, we extended our first assumption in such a way that we surmised that, in Spain, neither religious affiliation nor religious practice affects HRQoL through either health behaviours or lifestyles, and that they only exert an indirect effect on HRQoL through their impact on life satisfaction. It is worth noting that these assumptions were also supported by empirical evidence stemming from the Latinobaromenter 2001 –a nationwide survey conducted by the Spanish Centre for Sociological Research

[9] Compared with the general population, it has been found that Mormons and Seventh Day Adventists have lower incidence and mortality rates of cancers linked to tobacco and alcohol use (Grundmann, 1992; Fraser, 1999). Miller *et al.* (2000) also found, using data from a representative sample of American adolescents, that affiliation with conservative denominations was inversely associated with alcohol and illicit drug use.

(2001).This dataset gathers information about life satisfaction and self-reported health among a representative sample of the Spanish adult population (both variables are assessed using a five point single-item scale), although it lacks data on diagnosed chronic diseases. It also collects information regarding religious affiliation and practice. We used this dataset to examine whether religion is significantly associated with life satisfaction and self-reported health in Spain. To that end, we considered similar specifications to those presented in Table 2.3 –these estimates are available upon request to the authors. Our findings actually confirmed that neither religious affiliation nor religious practice is significantly associated with self-perceived health in Spain. On the contrary, we found a positive and relevant association between religious practice and life satisfaction.

On the other hand, our second identifying (i.e. that most chronic diseases only affect life satisfaction through their direct effect on HRQoL) is supported by some studies where SWB has been found to be more strongly related to self-perceived health than to objective measures of health (George and Landerman, 1984; Okun and George, 1984; Böckerman *et al.*, 2011). For instance, in a study about the effect of health on life satisfaction, using a representative sample of the Finnish adult general population, Böckerman *et al*. (2011) observed that, after controlling for HRQoL (as assessed by the EQ-5D index), only two of the eight chronic diseases considered were still significantly associated with lower levels of life satisfaction. Psychiatric disorders remained the health problems with the largest negative effect on life satisfaction. In this respect, it must be noted that we reached the same conclusion from Table 2.3 for the indicator of mental disorders. On this basis, we allowed mental disorders and poor circulation –whose coefficients were significantly different from zero in our estimates– to have a direct effect on life satisfaction. Remarkably, all our estimation results remain largely unchanged if we assume that no chronic disease has a direct effect on life satisfaction.

Table 2.5 displays the 3SLS estimates for the simultaneous equations system with the aforementioned exclusion restrictions.[10] As in Table 2.3, we standardised the coefficients and standard errors and we used corrective weights to recover the distribution of educational levels in the Spanish general population. We found life satisfaction and the SF-6D index to be significant determinants of each other at conventional significance levels. This implies that life satisfaction and HRQoL are simultaneously related. Consequently, the OLS estimates presented in Table 2.3 are biased and inconsistent. The comparison between these estimates and those in

[10] 2SLS estimates are similar to 3SLS ones. The exclusion restrictions passed tests of overidentificacion restrictions and both residuals passed Jarque-Bera normality tests.

Table 2.5 shows that the unidirectional approach severely underestimates the effect of life satisfaction on HRQoL. In particular, the standardised coefficient of the effect of life satisfaction on the SF-6D index presented in Table 2.5 (0.066) is 22.2% larger than the corresponding OLS estimate in Table 2.3 (0.054). The OLS model also underestimates the effect of the SF-6D index on life satisfaction, although to a lesser extent: this effect is 6.8% larger in the simultaneous equations model (0.079) than in the OLS model (0.074). As a result, whereas the OLS estimates show that the effect of life satisfaction on the SF-6D index is 27% lower than the opposite direction effect, our simultaneous equations estimates indicate that it is 16.5% lower. Therefore, the unidirectional approach underestimates the effect of life satisfaction on the SF-6D index as well as the opposite direction effect (by 18.2% and 6.3%, respectively). To get a sense of the magnitude of the simultaneous estimates we also calculated elasticities at the mean value of the variables. We found that a one percent increase in the SWLS summary score gives rise to a 0.297 percent increase in the SF-6D index. The elasticity for the opposite direction effect amounts to 0.316. Both elasticities are higher than those derived from the unidirectional approach, especially the elasticity for the effect of life satisfaction on the SF-6D index.

Table 2.5. Simultaneous equations approach. 3SLS estimates[a,b]

	SF-6D index		SWLS summary score	
	(1)		(2)	
	Coeff.	SE	Coeff.	SE
Constant	0.416**	0.162	24.704***	3.029
SWLS summary score	0.066*	0.036		
SF-6D index			0.079***	0.024
Age	0.213***	0.051	-0.181***	0.055
$Age^2/100$	-0.224***	0.051	0.203***	0.056
Female	-0.008	0.007	0.011	0.007
Single	0.028***	0.010	-0.036***	0.010
Divorced	0.004	0.008	-0.030***	0.007
Widowed	-0.015**	0.008	-0.012	0.008
Illiterate	-0.010	0.012	-0.041***	0.010
Low	0.005	0.008	-0.030***	0.010
Middle	0.005	0.008	-0.029***	0.009
Unemployed	0.005	0.007	-0.011	0.007
Homemaker	0.012	0.008	-0.006	0.009
Student	0.012	0.008	0.004	0.009
Retired	-0.003	0.011	-0.002	0.011

Table 2.5 (continued)

	SF-6D index		SWLS summary score	
	(1)		(2)	
	Coeff.	SE	Coeff.	SE
Less than €900	-0.008	0.012	-0.026**	0.010
€901-1,500	0.010	0.011	-0.038***	0.009
€1,501-2,000	-0.003	0.007	-0.003	0.008
Rural area	0.001	0.007	0.004	0.007
Realistic	-0.007	0.006	0.008	0.007
Pessimistic	-0.008	0.008	-0.021***	0.007
Non-practicing believer			0.029***	0.008
Practicing believer			0.031***	0.009
Allergy	-0.013**	0.006		
Arthritis/rheumatism	-0.026***	0.008		
Bronchitis	-0.019***	0.006		
Diabetes	-0.007	0.007		
Migraine	-0.013**	0.006		
Prostate	-0.000	0.006		
Musculoskeletal	-0.039***	0.007		
Poor circulation	-0.020***	0.008	0.018**	0.008
Cardiovascular	-0.019***	0.007		
Mental disorders	-0.062***	0.011	-0.027**	0.011
Hypertension	-0.016**	0.007		
Cholesterol	-0.003	0.006		
N	870		870	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Reference categories for discrete variables are: male; married/cohabiting; high educational level; employed/self-employed; more than €2,000 per month; optimistic; non-believer; with no chronic disease.

[a] The standardised coefficients and standard errors (SEs) were obtained by multiplying the original estimates by the standard deviation of the corresponding explanatory variable and dividing by the mean of the dependent variable.

[b] All regressions were estimated using corrective weights to change the distribution of educational attainment in the estimation sample for that in the Spanish adult general population.

Additionally, we analysed the potential heterogeneity in the relationship between life satisfaction and the SF-6D index by separately estimating the simultaneous equations model for men, women and for three age groups: ‘younger’ (under 36 years), ‘middle-aged’ (36 to 50 years) and ‘older’ respondents (over 50 years). This classification ensures that each of the three age groups accounts for at least 30% of the whole sample. In all cases, we used the same two exclusion restrictions as in the unconditional analysis. The estimation results are summarised in Table 2.6.[11] On the one hand, we found a simultaneous relationship between life satisfaction and the SF-6D index in the female group, but not in the male group. In the latter case, only the effect of the SF-6D index on life satisfaction is significant at conventional significance levels. For women, our estimates indicate that the effect of life satisfaction on the SF-6D index is around 23% larger than the opposite direction effect. Moreover, the estimated effect of the SF-6D index on life satisfaction for women is 50% larger than that estimated for men. On the other hand, we obtained evidence of simultaneity for younger and older respondents. Surprisingly, in both groups the effect of life satisfaction on the SF-6D index is larger than the reverse effect, especially in the latter group. For the middle-aged, only the effect of the SF-6D index on life satisfaction is significantly different from zero at conventional significance level. The latter effect is close in magnitude to that for younger respondents.

Table 2.6. Simultaneous equations approach by gender and age groups. 3SLS estimates[a]

			Dependent variables			
			SF-6D index		SWLS summary score	
			(1)		(2)	
Group	*N*	Explanatory variables[b]	Coeff.	SE	Coeff.	SE
All	870	SF-6D index			0.079***	0.024
		SWLS summary score	0.066*	0.036		
Men	434	SF-6D index			0.063***	0.023
		SWLS summary score	0.050	0.049		
Women	436	SF-6D index			0.095**	0.043
		SWLS summary score	0.117**	0.048		
Younger	339	SF-6D index			0.072**	0.032
		SWLS summary score	0.115**	0.046		
Middle-aged	239	SF-6D index			0.071***	0.026
		SWLS summary score	0.005	0.044		
Older	292	SF-6D index			0.102***	0.034
		SWLS summary score	0.178**	0.088		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[a] The coefficients and standard errors (SEs) (in parenthesis) were standardised in the same way as those displayed in Tables 2.3 and 2.5.
[b] The full list of regressors is the same as that displayed in Table 2.5.

[11] The estimated coefficients for the control variables are available upon request to the authors.

2.4. DISCUSSION

The study reported in this chapter has made several contributions to the scientific literature on the relationship between life satisfaction and self-perceived health. First, it adds to the incipient studies that have investigated the association between subjective well-being (SWB) and a preference-based measure of health-related quality of life (HRQoL) (Bobinac *et al.*, 2010; Böckerman *et al.*, 2011; Graham *et al.*, 2011). Indeed, this study has been the first one within this literature to use the SF-6D index for the assessment of HRQoL.[12] Moreover, there is no preceding research on the association between SWB and a preference-based measure of HRQoL using Spanish data. In this regard, it is worth pointing out that the SF-6D scores that we used to assess HRQoL were calculated using the Spanish algorithm for the SF-6D, which was derived from a representative sample of the Spanish adult general public that resembled the characteristics of the participants in the present study. Our main contribution, however, is that, unlike previous studies in which the relationship between SWB and HRQoL has been analysed under a unidirectional approach, we have explicitly accounted for the simultaneous relationship between both constructs. This is a relevant contribution because we have found that the unobserved factors that determine HRQoL and life satisfaction are significantly related, and that the fact of ignoring the simultaneous relationship between both variables results in biased and inconsistent estimates. Specifically, the conventional unidirectional approach underestimates the effect of life satisfaction on HRQoL as well as the opposite direction effect. Particularly relevant is the underestimation of the former effect (which amounts to18.2%), whereas the latter effect is underestimated to a lesser degree (by 6.3%). Either controlling for simultaneity or not, the effect of HRQoL on life satisfaction is higher than the reverse direction effect. However, after controlling for simultaneity, the differential between the two effects of interest lowers, although it is still large (approximately 20%). Thus, our results call for caution when interpreting estimates from regression models that do not take into account the simultaneous relationship between self-perceived health and satisfaction with life.

Furthermore, our results have also revealed that the association between life satisfaction and HRQoL is heterogeneous in terms of individual characteristics such as age and sex. We have obtained evidence of simultaneity for women, younger and older individuals. In all these three groups the effect of life satisfaction on HRQoL is larger than the opposite direction effect.

[12] Some recent studies have analysed the relationship between SWB (as outcome variable) and the dimensions and levels that comprise the SF-6D descriptive system (as explanatory variables) (Dolan, 2011; Mukuria and Brazier, 2013).

Conversely, in the male and middle-aged groups, only the effect of HRQoL on life satisfaction is significantly different from zero. Simultaneity studies conducted with larger datasets could further examine heterogeneity issues, not only in terms of age and sex, but also considering other variables. For instance, different groups could be distinguished according to HRQoL and life satisfaction scores.

Additionally, we have examined the determinants of both HRQoL and life satisfaction. Our estimates have shown that HRQoL is mainly influenced by chronic diseases, especially by mental disorders. It is also significantly associated with some individual characteristics (particularly with age and marital status). The same findings have been reported in preceding research on the determinants of HRQoL (Kind *et al.*, 1998; Franks *et al.*, 2003; Lubetkin *et al.*, 2005; Ferreira *et al.*, 2008). By contrast, we have found no relationship between HRQoL and factors such as gender, education, income and employment status. These results differ from those obtained in the aforementioned studies, where better HRQoL has been found to be associated with higher educational and income levels, whereas women and unemployed people have reported poorer HRQoL than men and those with a paid job, respectively. However, it is important to note that other studies using panel data have concluded that unemployment does not make an effect on self-reported health, because those with poor health are more likely to become unemployed (Böckerman and Ilmakunnas, 2009).

On the other hand, we have found that most demographic and socioeconomic factors (specifically age, marital status, income and unemployment) play a prominent role when explaining variations in life satisfaction responses, whilst chronic diseases are much less relevant in this case. In general, these findings are consistent with previous work on the correlates of life satisfaction (Di Tella *et al.*, 2001; Helliwell, 2003; Blanchflower and Oswald, 2004; Layard, 2005; Dolan *et al.*, 2008). With regard to gender, our results have not shown significant differences in average levels of life satisfaction between men and women. Albeit some researchers have found women to be more satisfied than men (Alesina *et al.*, 2004), the significant relationship between gender and life satisfaction seems to disappear when considering specific subgroups, such as informal carers (Van den Berg and Ferrer-i-Carbonell, 2007) or disabled people (Oswald and Powdthavee, 2008). Aside from these individual characteristics, we have found that life satisfaction is negatively related to pessimism and positively associated with religion, especially with the fact of being a practicing believer. These results are in line with preceding research on the role of personality (DeNeve and Cooper, 1998) and religious beliefs (Helliwell and Putnam, 2004) when explaining SWB. Conversely, as in other studies (George and Landerman, 1984; Okun and George, 1984), we have found that life satisfaction is weakly correlated with objective measures of health (i.e. chronic diseases). A possible explanation for this weak correlation is the adaptation to health problems (Brickman *et al.*, 1978).

Our study has a number of limitations, which are pointed out below, alongside some suggestions for further research. First, instead of measuring the three components of SWB (life satisfaction, positive affect and negative affect), we have restricted our research to the cognitive component of SWB (i.e. life satisfaction). The same choice has been made by many other researchers before us, but we are aware that a promising line for future research would be to include the affective component of SWB in further simultaneity studies.

Second, respondents in this study were asked if they had any diagnosed chronic disease, but not about severity. Given that the effect of a particular health state on life satisfaction and HRQoL is likely to differ according to its degree of severity, future studies should control not only for the presence of chronic conditions, but also for how severe they are. By the same token, other studies might focus on the role of adaptation to chronic diseases in explaining the association of life satisfaction and self-perceived health with objective measures of health.

Third, we have used two specific instruments –namely, the utility-based SF-6D index and the Satisfaction With Life Scale (SWLS)– in order to make HRQoL and life satisfaction operative concepts. In consequence, our estimates are only valid with reference to these specific measures. Therefore, the use of other utility-based measures of HRQoL (e.g. the EQ-5D index) could lead to different estimates. This potential variability is an intrinsic feature of the existing preference-based health state classification systems, in such a way that it is clear that ‘utilities from different systems are not interchangeable’ (Stiggelbout, 2006). In addition, there is no absolute guideline to prefer one instrument over another, ‘so often the choice will be a pragmatic one’. We chose the SF-6D because it allows the calculation of utilities from SF-36 responses, one of the most widely evaluated measures of HRQoL in the world (Garratt *et al.*, 2002). Furthermore, it is particularly suitable to detect small health changes at the upper end of the distribution (Grieve *et al.*, 2009), which is especially important when using samples drawn from the general population. At the top of these arguments there is also the availability of a recent scoring algorithm for the SF-6D estimated for the Spanish general public (Abellán *et al.*, 2012), which lowers the ‘floor’ of the SF-6D range of utilities, one of the empirical limitations of this instrument. We would like to emphasise that, although it would be undoubtedly very interesting to expand the analysis of the simultaneous relationship between life satisfaction and HRQoL (or, more generally, between SWB and self-perceived health) to other settings and measures, our estimates are valid for the Spanish general public, as previously shown.

Another important concern that deserves further investigation is to address the simultaneity issue using panel data techniques, which would reinforce the causal content of the estimates by, for example, controlling for unobserved time-invariant individual characteristics and/or by measuring the explanatory variables in the time period preceding that of the dependent variable (Ferrer-i-Carbonell and Frijters, 2004). It is worth pointing out that having repeated observations from a sample of individuals over time periods is not a necessary condition for a simultaneous equations model to be identified. Indeed, the identifying conditions of a simultaneous equations model (i.e. the rank and order conditions) were fulfilled in our application, by means of our exclusion restrictions. Moreover, we controlled for some relatively persistent individual characteristics, such as personality traits, chronic diseases and religious beliefs, which are likely to be uncorrelated with self-perceived health, income and other shocks that determine current HRQoL and life satisfaction and, thus, provide confidence on the robustness of our estimates.

Several implications can be drawn from our results. First, our results suggest that the unidirectional estimates obtained in previous studies on the relationship between life satisfaction and self-perceived health are potentially biased and inconsistent. Consequently, if unidirectional estimates are used to predict the effect of life satisfaction on self-perceived health (or the reverse effect), decisions based on those predictions will be wrong. Thus, our study calls for using simultaneous equations estimates to get optimal policy decisions. Remarkably, this holds regardless of whether the benefits of the policy under evaluation are assessed only in terms of its effect on HRQoL or also on life satisfaction. In this respect, it would be interesting to re-estimate previous unidirectional estimates on the relationship between life satisfaction and self-perceived health using a simultaneous equations scheme in order to ascertain if they are also biased. Likewise, future research based on new data should consider the interplay between life satisfaction and HRQoL, instead of treating them as independent variables.

Second, the fact that, according to our estimates, the impact of HRQoL on life satisfaction is larger than the opposite direction effect, suggests that health care interventions can be useful tools for the improvement of SWB, which should be ultimately one of the broad goals of public policies. In the reverse way, non-health related policies (e.g. educational and employment policies) that can potentially increase life satisfaction may cause positive side effects on HRQoL, which, in turn, would enhance satisfaction with life.

Finally, the results reported in this chapter might also have implications for the economic evaluation of health care programmes. For instance, at present it is not clear which kind of programmes –preventive or curative ones– should be prioritised given the same gain in HRQoL (or in QALYs). Some researchers (Johannesson and Johansson, 1997; Ubel, 1999) have found empirical support for preventive practices, whereas others (Corso *et al.*, 2002; Schwappach, 2003) have obtained opposite evidence. If preventive interventions (for instance, promotion of healthy lifestyles) have a larger direct effect on people's life satisfaction than curative treatments, and we accept the fulfillment of SWB as a goal of health policies, then we would have a rationale for giving a greater weight to preventive interventions over curative ones.

2.5. APPENDIX 2A

The questionnaire

PRESENTACIÓN

La presente encuesta forma parte de un estudio científico, realizado por investigadores de la Universidad de Murcia pertenecientes al Grupo de Trabajo en Economía de la Salud (GTES), con el apoyo financiero de la Dirección General de Planificación, Ordenación Sanitaria y Farmacéutica e Investigación de la Consejería de Sanidad y Consumo de la Comunidad Autónoma de la Región de Murcia.

El estudio tiene como finalidad conocer con mayor precisión el estado de salud de la población murciana. Asimismo se pretende investigar en qué medida el estado de salud influye en el grado de satisfacción que se siente con la propia vida. El cuestionario es completamente anónimo y las respuestas que dé usted a las preguntas que en él se formulan son totalmente confidenciales, y sólo serán objeto de análisis con fines investigadores. Le agradecemos mucho su colaboración, pues sus respuestas servirán para informar con mayor exactitud a los poderes públicos de la Región de Murcia acerca de cuáles son los problemas de salud más importantes en la actualidad.

PARTE 1: PREGUNTAS SOBRE SATISFACCIÓN CON LA VIDA

1. A continuación se presentan cinco afirmaciones con las que usted puede estar de acuerdo o en desacuerdo. Utilizando la siguiente escala de 1 a 7, en la que el 1 significa "*Completamente en desacuerdo*" y el 7 "*Completamente de acuerdo*", indique su grado de acuerdo con cada una de las afirmaciones escogiendo el número apropiado. Por favor, responda a las preguntas abierta y sinceramente.

		Completamente en desacuerdo	En desacuerdo	Más bien en desacuerdo	Ni de acuerdo ni en desacuerdo	Más bien de acuerdo	De acuerdo	Completamente de acuerdo
a)	En la mayoría de las cosas, mi vida está cerca de mi ideal.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$
b)	Las condiciones de mi vida son excelentes.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$
c)	Estoy satisfecho con mi vida.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$
d)	Hasta ahora, he conseguido las cosas que para mí son importantes en la vida.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$
e)	Si volviese a nacer, no cambiaría casi nada de mi vida.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$
f)	En general, soy feliz.	□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$

PARTE 2: PREGUNTAS RELATIVAS A SU ESTADO DE SALUD

INSTRUMENTO SF-36

2. Para cada una de las siguientes preguntas, por favor marque con una cruz la casilla que mejor corresponda a su respuesta.

2.1. En general, usted diría que su salud es:

Excelente	Muy buena	Buena	Regular	Mala
□1	□2	□3	□4	□5

2.2. ¿Cómo diría usted que es su salud actual, comparada con la de hace un año?:

Mucho mejor ahora que hace un año	Algo mejor ahora que hace un año	Más o menos igual que hace un año	Algo peor ahora que hace un año	Mucho peor ahora que hace un año
□1	□2	□3	□4	□5

2.3. Las siguientes preguntas se refieren a actividades o cosas que usted podría hacer en un día normal. Su salud actual, ¿le limita para hacer esas actividades o cosas? Si es así, ¿cuánto?

	Sí, me limita mucho	Sí, me limita un poco	No, no me limita nada
a) Esfuerzos intensos, tales como correr, levantar objetos pesados, o participar en deportes agotadores.	□1	□2	□3
b) Esfuerzos moderados, como mover una mesa, pasar la aspiradora, jugar a los bolos o caminar más de 1 hora.	□1	□2	□3
c) Coger o llevar la bolsa de la compra.	□1	□2	□3
d) Subir varios pisos por la escalera.	□1	□2	□3
e) Subir un solo piso por la escalera.	□1	□2	□3
f) Agacharse o arrodillarse.	□1	□2	□3

g) Caminar un kilómetro o más.	□1	□2	□3
h) Caminar varios centenares de metros.	□1	□2	□3
i) Caminar unos 100 metros.	□1	□2	□3
j) Bañarse o vestirse por sí mismo.	□1	□2	□3

2.4. Durante las 4 últimas semanas, ¿con qué frecuencia ha tenido alguno de los siguientes problemas en su trabajo o en sus actividades cotidianas, a causa de su salud física?

	Siempre	Casi siempre	Algunas veces	Solo alguna vez	Nunca
a) ¿Tuvo que reducir el tiempo dedicado al trabajo o a sus actividades cotidianas?	□1	□2	□3	□4	□5
b) ¿Hizo menos de lo que hubiera querido hacer?	□1	□2	□3	□4	□5
c) ¿Tuvo que dejar de hacer algunas tareas en su trabajo o en sus actividades cotidianas?	□1	□2	□3	□4	□5
d) ¿Tuvo dificultad para hacer su trabajo o sus actividades cotidianas (por ejemplo, le costó más de lo normal?	□1	□2	□3	□4	□5

2.5. Durante las 4 últimas semanas, ¿con qué frecuencia ha tenido alguno de los siguientes problemas en su trabajo o en sus actividades cotidianas, a causa de algún problema emocional (como estar triste, deprimido o nervioso)?

	Siempre	Casi siempre	Algunas veces	Solo alguna vez	Nunca
a) ¿Tuvo que reducir el tiempo dedicado al trabajo o a sus actividades cotidianas por algún problema emocional?	□1	□2	□3	□4	□5
b) ¿Hizo menos de lo que hubiera querido hacer, por algún problema emocional?	□1	□2	□3	□4	□5

c) ¿Hizo su trabajo o sus actividades cotidianas menos cuidadosamente que de costumbre, por algún problema emocional?	□1	□2	□3	□4	□5

2.6. 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
□1	□2	□3	□4	□5

2.7. ¿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
□1	□2	□3	□4	□5	□6

2.8. 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
□1	□2	□3	□4	□5

2.9. 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	Solo alguna vez	Nunca
a) se sintió lleno de vitalidad?	□1	□2	□3	□4	□5
b) estuvo muy nervioso?	□1	□2	□3	□4	□5
c) se sintió tan bajo de moral que nada podía animarle?	□1	□2	□3	□4	□5
d) se sintió calmado y tranquilo?	□1	□2	□3	□4	□5
e) tuvo mucha energía?	□1	□2	□3	□4	□5

f) se sintió desanimado y deprimido?	□1	□2	□3	□4	□5
g) se sintió agotado?	□1	□2	□3	□4	□5
h) se sintió feliz?	□1	□2	□3	□4	□5
i) se sintió cansado?	□1	□2	□3	□4	□5

2.10. Durante las 4 últimas semanas ¿con qué frecuencia la salud física o los problemas emocionales le han dificultado sus actividades sociales (como visitar a los amigos o familiares)?

Siempre	Casi siempre	Algunas veces	Solo alguna vez	Nunca
□1	□2	□3	□4	□5

2.11. Por favor, diga si le parece CIERTA o FALSA cada una de las siguientes frases:

	Totalmente cierta	Bastante cierta	No lo sé	Bastante falsa	Totalmente falsa
a) Creo que me pongo enfermo más fácilmente que otras personas	□1	□2	□3	□4	□5
b) Estoy tan sano como cualquiera	□1	□2	□3	□4	□5
c) Creo que mi salud va a empeorar	□1	□2	□3	□4	□5
d) Mi salud es excelente	□1	□2	□3	□4	□5

3. ¿Le ha dicho un médico que usted padece actualmente alguno de los problemas de salud siguientes?

	Sí	No
a) Alergias crónicas	□1	□2
b) Artritis o reumatismo	□1	□2
c) Bronquitis crónica	□1	□2
d) Diabetes	□1	□2
e) Migrañas/Jaquecas	□1	□2
f) **(SOLO HOMBRES)** Trastornos de próstata	□1	□2
g) Dolor o molestias de espalda, cuello, hombro, cintura	□1	□2
h) Mala circulación	□1	□2
i) Trastornos cardiacos	□1	□2
j) Problemas nerviosos/depresiones	□1	□2
k) Varices en las piernas	□1	□2
l) Hipertensión	□1	□2
m) Colesterol	□1	□2
n) (SOLO MUJERES) Dolor menstrual	□1	□2
o) Otras enfermedades crónicas (especificar, máximo dos)		

PARTE 3: PREGUNTAS SOBRE CARACTERÍSTICAS PERSONALES

4. En términos generales, usted se considera una persona:

Optimista	Realista	Pesimista
□1	□2	□3

5. En cuanto a las creencias religiosas, usted se considera:

No creyente	Creyente, aunque no practicante	Creyente practicante
□1	□2	□3

6. ¿Pertenece usted a alguna institución relacionada con alguna confesión religiosa (asociación, ONG, grupo de oración, etc.)?

Sí	No
□1	□2

7. Marque o rellene según corresponda en las siguientes cuestiones.

a) Sexo

Hombre	Mujer
□1	□2

b) Edad …......... años

c) Lugar de residencia ……………………………………………

d) Estado civil

Soltero	Casado (o pareja de hecho)	Separado/ Divorciado	Viudo
□1	□2	□3	□4

	Sin estudios	Primarios (EGB o similar)	Secundarios (BUP, FP, COU)	Superiores (Diplomado, Licenciado)
e) Nivel de estudios terminados	□1	□2	□3	□4

	Empleado o Autónomo	Parado	Ama de Casa	Estudiante	Jubilado
f) Situación Laboral	□1	□2	□3	□4	□5

	Menos de 900 €	De 901 a 1.500 €	De 1.501 a 2.000 €	De 2.001 a 3.000 €	Más de 3.000 €
g) Nivel de ingresos mensuales	□1	□2	□3	□4	□5

i) Número de hijos que conviven con Ud. ……….

j) Número de personas, incluido Ud. mismo, que conviven en su hogar ………..

Comentarios/Sugerencias:

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

MUCHAS GRACIAS POR SU COLABORACIÓN

2.6. APPENDIX 2B

Table 2B.1. Dimensions and levels of the SF-6D descriptive system

Level	Dimension
	Physical Functioning
1	Your health does not limit you in vigorous activities
2	Your health limits you a little in vigorous activities
3	Your health limits you a little in moderate activities
4	Your health limits you a lot in moderate activities
5	Your health limits you a little in bathing and dressing
6	Your health limits you a lot in bathing and dressing
	Role limitations
1	You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems
2	You are limited in the kind of work or other activities as a result of your physical health
3	You accomplish less than you would like as a result of emotional problems
4	You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems usual activities (e.g. work, study, housework, family or leisure activities)
	Social functioning
1	Your health limits your social activities none of the time
2	Your health limits your social activities a little of the time
3	Your health limits your social activities some of the time
4	Your health limits your social activities most of the time
5	Your health limits your social activities all of the time
	Pain
1	You have no pain
2	You have pain but it does not interfere with your normal work (both outside the home and housework)
3	You have pain that interferes with your normal work (both outside the home and housework) a little bit
4	You have pain that interferes with your normal work (both outside the home and housework) moderately
5	You have pain that interferes with your normal work (both outside the home and housework) quite a bit
6	You have pain that interferes with your normal work (both outside the home and housework) extremely

Table 2B.1 (continued)

Level	Dimension
	Mental health
1	You feel tense or downhearted and low none of the time
2	You feel tense or downhearted and low a little of the time
3	You feel tense or downhearted and low some of the time
4	You feel tense or downhearted and low most of the time
5	You feel tense or downhearted and low all of the time
	Vitality
1	You have a lot of energy all of the time
2	You have a lot of energy most of the time
3	You have a lot of energy some of the time
4	You have a lot of energy a little of the time
5	You have a lot of energy none of the time

CHAPTER 3

MONETARY VALUATION OF INFORMAL CARE BASED ON CARERS' AND NON-CARERS' PREFERENCES

ABSTRACT

This chapter reports the results of a willingness to accept (WTA) study for the monetary valuation of informal care based on the preferences of two groups: informal carers (n = 202) and non-carers (n = 200). In both samples, three WTA questions were posed in reference to a hypothetical caring scenario: WTA for one extra hour of care in general terms, WTA for one additional hour performing the most unpleasant task, and WTA for one extra hour undertaking the least unpleasant task. Furthermore, informal carers were also asked for the compensation they would require if they had to be in charge of their actual care recipient for one extra hour. No significant differences were found between the distributions of carers' and non-carers' WTA values, with carers' and non-carers' mean/median WTA values (in general terms) amounting to €6.4/€5.5 and €6.5/€5.5, respectively. Overall, respondents' valuations were sensitive to and consistent with their preferences over the tasks to be carried out in the extra hour of informal care, although a considerable proportion of subjects (around 25% in each group) stated the same value in the three hypothetical WTA questions. On average, carers' required a lower monetary compensation for one extra hour taking care of their loved one (mean/median WTA values €5.2/€4.5) than if they had to devote that time to look after the hypothetical care recipient. However, more than half of carers stated the same value both under their actual situation and under the hypothetical scenario, suggesting that carers' valuations were influenced by their own experience providing informal care.

3.1. INTRODUCTION

Informal carers play a key role in the health care sector since they provide a great part of the assistance that their ill loved ones (i.e. relatives, friends or neighbours) need to perform the activities of daily living (Norton, 2000). Not surprisingly, nationwide estimates of the economic value of informal care far exceed those of formal (or professional) care (Arno *et al.*, 1999; Buckner *et al.*, 2011). The provision of informal care may cause a profound impact on the different dimensions of carers' lives. For instance, informal carers can bear substantial opportunity costs due to the time they have to give up in other activities, such as a paid job, family duties, social relationships or leisure time (Hepburn and Barling, 1996; Lilly *et al.*, 2007). Furthermore, as a vast body of evidence has shown, informal carers may undergo great burden, physical and psychological problems, and even increased risk of mortality (Montgomery *et al.*, 1985; Schulz *et al.*, 1997; Vitaliano *et al.*, 2004; Hirst, 2005; Coe and Van Houtven, 2009). Nevertheless, informal carers can also experience positive feelings (e.g. fulfillment, reward, gain, satisfaction) as a result of the care they provide (Cohen *et al.*, 2002; Brouwer *et al.*, 2005; Al-Janabi *et al.*, 2010).

Despite the significant costs and effects resulting from the provision of informal care, economic evaluations of health care interventions usually ignore them (Goodrich *et al.*, 2012), which is equivalent to view it as a costless resource and, therefore, as a free substitute for professional care (Koopmanschap *et al.*, 2008). This neglect can lead to make wrong resource allocation decisions (Van den Berg *et al.*, 2005c), for example, by favouring those treatments that rely heavily on informal care. It is worth noting that economic evaluations in which informal care is not accounted for contravene the commonly recommended societal perspective, insofar it claims that all relevant costs and effects stemming from the intervention under study should be considered, regardless of who experiences them (Russell *et al.*, 1996; Drummond *et al.*, 2005). Even in countries where the societal perspective is not adopted and a narrow health care system perspective is taken instead (e.g. in England and Wales), it has been advocated that economic evaluations incorporate health effects on carers, whenever they are clinically relevant (National Institute for Health and Clinical Excellence, 2007, 2013).

The frequent exclusion of informal care in economic evaluations may be related to the lack of consensus about how best to account for it in these studies, along with the absence of standardised methods that are able to measure and value informal care in a feasible manner (Van den Berg *et al.*, 2004). Given that, to a large extent, the costs of informal care depend on the time invested in caring, it has been suggested that the change in informal care time should be included in the cost side of economic evaluations, as a direct non-medical cost (Luce *et al.*, 1996). The monetary value of this time input could be used in all types of economic evaluations (i.e. cost-benefit, cost-effectiveness and cost-utility analyses) (Koopmanschap *et al.*, 2008).

There are several methods that can be used to derive a monetary value for informal care. Traditionally, it has been recommended that the time spent on providing informal care be monetised using either the opportunity cost method or the proxy good method (also known as replacement cost method) (Luce *et al.*, 1996; Posnett and Jan, 1996; Drummond *et al.*, 2005), which are both considered to be revealed preference methods.

In the opportunity cost method (Liu *et al.*, 2002; Van den Berg *et al.*, 2006), informal care is valued at the opportunity cost attributable to the provision of informal care. This opportunity cost is estimated as the value of the benefits that the informal carer relinquishes because of the time devoted to informal care. In general, this value is approximated by an individual's market wage rate. If the informal carer is unemployed, a proxy is used for this wage rate, for example, the reservation wage rate (i.e. the wage rate for which the informal carer is willing to supply at least one hour of paid labour) (Koopmanschap *et al.*, 2008). A disadvantage of the opportunity cost method is that a same hour of informal care is valued differently depending on the carer's wage rate (e.g. depending on his/her job).

The proxy good method (Oliva and Osuna, 2009; Sabes-Figuera *et al.*, 2010) values the time spent on informal caring at the price of a close market substitute (e.g. professional home care). Different values can be obtained depending on the type of caring task. For example, housework is valued at the market wage rate of a professional housekeeper, personal care is valued at the market wage rate of a professional nurse, etc. This method is quite limited because it assumes that there are no differences in quality between informal care and professional care (McDaid, 2001), and that informal carers and care recipients are indifferent between the two alternatives of care, which are two unrealistic presumptions (Van den Berg *et al.*, 2005b). In summary, both the opportunity cost method and the proxy good method value exclusively the costs associated with the time devoted to provide informal care, instead of the full impact that caring has on the carer's life, which can entail both positive and negative aspects (Brouwer *et al.*, 2005; Al-Janabi *et al.*, 2008). Furthermore, neither of the two methods accurately assesses carers' and care recipients' preferences. To overcome these limitations, alternative methods have been proposed and applied for the monetary valuation of informal care. These include the contingent valuation (CV) method (Van den Berg *et al.*, 2005c; Gustavsson *et al.*, 2010), multi-attribute stated preference methods –conjoint analysis (CA) (Van den Berg *et al.*, 2005a; Van den Berg *et al.*, 2008) and discrete choice experiments (DCEs) (Mentzakis *et al.*, 2010)– and the well-being valuation (WBV) method (Van den Berg and Ferrer-i-Carbonell, 2007).

CV is a widely used technique for the valuation of non-market goods, such as environmental commodities (Cummings *et al.*, 1986; Haab and McConnell, 2002) and health outcomes (Olsen *et al.*, 2005; Donaldson *et al.*, 2012). Although still scarce, a number of incipient studies have used this method for the monetary valuation of informal care (Van den Berg *et al.*, 2005b; Van den Berg *et al.*, 2005c; De Meijer *et al.*, 2010; Gustavsson *et al.*, 2010). The CV method can be applied by asking individuals about their maximum willingness to pay (WTP) in order to obtain a potential benefit (e.g. for providing one fewer hour of informal care) or, alternatively, about their minimum willingness to accept (WTA) as a compensation for a potential loss (e.g. for providing one additional hour of informal care). According to the standard economic theory, WTP and WTA values for a same good should be fairly close (Wilig, 1976), unless the good represents a substantial proportion of subjects' income and the associated transaction costs are large (Brown and Gregory, 1999). Notwithstanding, the typical finding in CV studies is that WTA significantly surpasses WTP (Knetsch and Sinden, 1984), which can be due to a variety of reasons –see Brown and Gregory (1999) for a detailed explanation. Contrary to this general pattern, those studies that have elicited both WTA and WTP values for informal care have found small differences between the two values, with WTA being slightly higher than WTP (Van den Berg *et al.*, 2005b; De Meijer *et al.*, 2010). Van den Berg *et al.* (2005b) speculated that the small discrepancy they found between carers' WTA and WTP values for informal care might be explained because informal carers seem to have relatively well-defined preferences for informal care and are used to make decisions regarding this 'commodity'. Although the NOAA Panel favoured the use of WTP over WTA for the monetary valuation of non-market goods (Arrow *et al.*, 1993), it has been argued that it is more appropriate to use WTA instead of WTP when subjects are asked to value a potential welfare loss, while WTP is more suitable when subjects have to value a potential welfare gain (Bromley, 1995). Given that the provision of informal care entails a sacrifice (in terms of the time that is given up in other activities), it has been stated that it is more appropriate and natural to value informal care using WTA rather than WTP (Van den Berg *et al.*, 2005c), although there is not a formal consensus on this respect.

CV has been found to be sensitive to the carers' characteristics and, allegedly, it reflects their preferences (Van den Berg *et al.*, 2005c; De Meijer *et al.*, 2010). However, a drawback of this method is that it mainly focuses on money and some informal carers may find it difficult and even unethical to indicate how much money they would be willing to pay (or to accept) to spent less (or more) time on assisting a loved one –a similar problem can arise when CV is applied with care recipients. As a result, CV may lead to strategic or protest answers (Van den Berg *et al.*, 2005a). Moreover, CV studies can result in different types of biases and anchoring effects (Van Exel *et al.*, 2006). In part, these problems could be avoided using either CA or DCEs. In both types of methods respondents are presented with a sequence of hypothetical scenarios, which are defined in terms of different attributes and levels. If one of the attributes is the price of the alternative, it is possible to derive a monetary value for it. In CA studies respondents are generally asked to rank or to rate the different scenarios presented, whilst DCEs entail maiking choices between scenarios. Compared with the CV method, both CA and DCEs have a number of advantages (Van den Berg *et al.*, 2005a). For instance, in these techniques respondents are not directly asked to state a monetary value, but to make trade-offs between different aspects of the presented scenarios, which can lower the risk of strategic and protest answers[1]. Moreover, they provide more information about respondents' preferences and they can be better able to reflect the heterogeneity of informal care because they do not focus only on money. Nonetheless, one of the main limitations of conventional CA and DCEs is that they do not provide valuations at an individual level (Lancsar *et al.*, 2013). In addition, these methods can be cognitively demanding, since respondents have to assess different scenarios simultaneously. As a result, they can result in low response rates. For example, in a CA study for the monetary valuation of informal care conducted by Van den Berg *et al*. (2008), only 26.6% of respondents completed the questionnaire, and the response rate obtained in a DCE study undertaken by Mentzakis *et al.* (2010) was even lower (20%).

Whilst CV, CA and DCEs are stated preference methods (i.e. they are intended to elicit respondents' preferences using survey questions), the WBV method aims to value the full impact (costs and benefits) of informal care on carers' well-being. In this method, the value assigned to an hour of informal care represents the monetary compensation needed to maintain the same level of informal carer's well-being after providing one extra hour of informal care. This value is calculated as the trade-off between income and the number of hours of informal care provided so as to maintain informal carer's well-being unchanged. Van den Berg and Ferrer-i-Carbonell (2007) found lower values for one additional hour of informal care with the WBV method than with the WTA technique. According to these researchers, this result could be

[1] In CV studies, respondents provide a protest zero when they reject to pay anything (when asking for WTP) or to receive any compensation (when asking for WTA) but they actually ascribe a value different from zero to the 'good' under consideration. Respondents who provide a protest answer generally argue that the Government should pay the good being valued (in case of WTP questions) or that it is unethical to attach a monetary value to that good (Halstead *et al.*, 1992).

due to the fact that CV does not capture properly the positive aspects associated with the provision of informal care.

Apart from the choice of the valuation method, another key issue is to decide who should be asked to value informal care. When the WBV method is used, the answer is clear, because this technique uses data collected through survey questions which are directly posed to a sample of informal carers. By contrast, when a stated preference method is used, preferences for informal care can be elicited from various sources. In principle, the most straightforward way is to ask actual informal carers (Van den Berg *et al.*, 2004), because they are the individuals who directly experience the positive and negative effects of informal care and, thus, they are the best informed subjects. Furthermore, informal carers are used to make decisions concerning informal care (Van den Berg *et al.*, 2008). For that reason, stated preference studies on the valuation of informal care reported until now have focused on the carers' perspective (Van den Berg *et al.*, 2005c; Gustavsson *et al.*, 2010; Mentzakis *et al.*, 2010). Additionally, some CV studies have elicited monetary valuations for informal care from both informal carers' and care recipients' point of views (Van den Berg *et al.*, 2005b; De Meijer *et al.*, 2010), in an attempt to reflect the interdependency between the preferences of the two groups. Another alternative is to elicit preferences for informal care from the general public. Hitherto, however, no stated preference study has derived a monetary value for informal care using this approach. From a societal perspective, there are several reasons why it is important to consider public's preferences for informal care. First, the general population comprises potential, actual or former informal carers or care recipients (Van den Berg *et al.*, 2004). Second, they partly bear the indirect costs associated with informal care (e.g. the increase in consumption of health care resources if the carer suffers from health problems due to his/her caring duties). Moreover, in theory, members of the general public who are neither informal carers nor care recipients (who will be called 'non-carers' throughout this chapter) may be more objective (although presumably less informed) than informal carers and care recipients are (Van den Berg *et al.*, 2004), which can lower the risk of strategic or self-interested responses and the incidence of protest zeros. However, an objection to the valuation of informal care from the public's perspective is that non-carers may not have well defined (or complete) preferences for informal care (especially if they have no prior experience as informal carers) and, as a result, compared with informal carers, they could be more likely to provide no answer or to anchor in the cost of professional care if they were asked to place a monetary value on informal care. Consequently, non-carers' values for informal care may be less trustworthy than those elicited from a sample of informal carers.

In order to shed light on these issues, the primary objectives of the present study are to obtain a monetary value for informal care based on the stated preferences of a sample of non-carers and to compare this valuation with that elicited from a sample of actual informal carers ('carers' hereafter). To that end, individuals in both groups were asked to state the minimum monetary compensation (WTA) they would demand if they had to spend one extra hour per day taking care of a fictitious care recipient. Since some studies have found that the value of an hour of informal care depends on the activity to be carried out (Van den Berg *et al.*, 2005c; Mentzakis *et al.*, 2010), within each group, three WTA values were compared to test for logical consistency: WTA for one extra hour of care per day in general terms, WTA for one additional hour per day undertaking the activity ranked as the least preferred task, and WTA for one extra hour per day performing the activity regarded as the most preferred task. Finally, this study also examines whether carers' valuations in reference to a hypothetical scenario are influenced by their own caring situation, by comparing carers' WTA for one extra hour per day taking care of a hypothetical person with carer's WTA for one additional hour per day looking after their own care recipient.

The decision of using CV instead of a multi-attribute stated preference method was made taking into account that, as previously mentioned, these techniques seem to be more burdensome than CV, with the consequent risk of obtaining a lower response rate. On the other hand, since WTA is more suitable to value a potential welfare loss and WTP is more appropriate to value a potential gain (Bromley, 1995), we used the former method instead of the latter, given that the provision of informal care entails a sacrifice (e.g. in terms of the time that is not devoted to other activities).

3.2. METHODS

3.2.1. Participants

The data used in this study come from a survey that was specifically designed for the monetary valuation of informal care. The selection of participants and the data collection were carried out by eight trained interviewers, over a two-week period. The questionnaire was face-to-face administered by one of the interviewers at respondents' home (located throughout the Region of Murcia, Spain) and respondents received no reward of any kind. Prior to the final survey, a pilot study was conducted with a convenience sample (n = 66) of students and teachers at the University of Murcia.

Two independent samples were selected –one of informal carers and another of non-carers–, with a target size of 200 individuals in each group. The sample of non-carers was selected according to a quota system based on gender (50% of men and 50% of women) and age, so as to resemble the Spanish adult general population in terms of these characteristics. Potential respondents were first approached by telephone using random digit dialing. After explaining the aim of the study, the subjects were asked if they were willing to take part in the survey. Those who agreed to participate were then asked if they devoted at least one hour per week to assist an ill loved one (relative, friend or neighbour) with the activities of daily living. If the answer was 'yes', the individual was assigned to the sample of informal carers. If the answer was 'no', the subject was subsequently asked if he/she received at least one hour of informal care per week. Those who answered answered 'yes' were considered to be care recipients and were left out from the study. Those who answered 'no' were allocated to the sample of non-carers. In the process of selecting the 200 non-carers, we found 43 carers who were willing to participate in the survey. In order to achieve a balance between the two samples (in terms of size), it was necessary to increase the sample of informal carers. This was done by looking for informal carers in different settings (the neighbourhoods of the interviewers, primary care centres, hospitals and day care centres). The final sample of informal carers comprised 202 subjects. No quotas were used for this group.

3.2.2. The questionnaire

Two versions of the questionnaire were designed: one for carers and the other one for non-carers. Table 3.1 outlines the structure of the two types of questionnaires, while the full questionnaires are presented in Appendix 3A. The questionnaire for carers was larger than the questionnaire for non-carers, because the former contained two additional sets of questions: a variety of care-related questions and a question asking for carers' WTA under their own caring situation ('WTA_{own}'). Notwithstanding, the two versions had the same question order (in the shared sections).

The two types of questionnaires started with a set of three core questions intended to elicit respondents' WTA for one additional hour of care per day under the hypothetical scenario presented in Figure 3.1. This scenario described the health state of a fictitious disabled person in terms of the DEP-6D classification system (Rodríguez, 2012). This instrument is used to characterise dependency states, which is done by means of six dimensions (or attributes), each one with three or four levels of severity. The six attributes (and the number of levels) are: eat

(3), incontinence (3), personal care (4), mobility (4), housework (3) and mental health problems (4). For each dimension, level 1 represents the mildest level of dependency, whereas the upper level (3 or 4, depending on the dimension) stands for the highest level of dependency. By combining the attributes with the levels of severity, a total of 1,728 dependency states can be described. The selected dependency state was that coded as '334332'. Additionally, the hypothetical scenario also detailed the number of daily hours to be invested in four different types of caring tasks -assistance in personal care, mobility, housekeeping and practical activities- in order to satisfy the daily life needs of the hypothetical care recipient.

Table 3.1. Overview of the questionnaires

Questions	Carers	Non-carers
$WTA_{hypothetical}$		
• $WTA_{general}$: WTA for one additional hour of care per day (without specifying the task to be provided)	X	X
• Ranking of tasks described in the hypothetical scenario (from the least to the most preferred tasks)	X	X
• WTA_{worst}: WTA for one extra hour of care per day performing the least preferred task	X	X
• WTA_{best}: WTA for one extra hour of care per day performing the most preferred task	X	X
Care-related questions[a]	X	
WTA_{own}: WTA for one additional hour of care per day under the actual caring situation	X	
Prior experience as informal carer		X
Whether the non-carer knew some informal carer		X
Demographic, socioeconomic and health-related characteristics	X	X

[a] When the carer looked after more than one person, the care-related questions (and the WTA_{own} question) were posed in relation to the main care recipient.

The first WTA question (labelled as '$WTA_{general}$') was posed in general terms (i.e. without specifying the task to be performed in the extra hour of informal care). The full wording of this first question is also detailed at the bottom of Figure 3.1 ('Suppose you should look after that person for one additional hour every day…'). Given that the hypothetical situation entailed the provision of four types of activities, respondents were asked to rank them, from the least preferred (or the most unpleasant) one to the most preferred (or the least unpleasant) one. After this ranking exercise, two further WTA questions were posed, which were formulated in the same terms as the $WTA_{general}$ question, except for the fact that, instead of asking for the monetary compensation demanded for one additional hour of care (in general), it was specified that the extra hour of care should be devoted to the least preferred task (in the first question) and to the most preferred task (in the second question). Henceforth, WTA in relation to the least preferred task and to the most preferred task will be abbreviated as 'WTA_{worst}' and 'WTA_{best}', respectively, while the abbreviation '$WTA_{hypothetical}$' (as opposed to WTA_{own}) will be used when alluding to $WTA_{general}$, WTA_{worst} and WTA_{best} as a whole. Before formulating the $WTA_{general}$ question, carers were explicitly asked to abstract from their own caring situation when answering the three $WTA_{hypothetical}$ questions.

A payment card format was used in all WTA questions. The following amounts of money (in euros per day) were randomly offered to each respondent: €0, €1, €2, €3, €4, €5, €6, €8, €10, €12 and €15.[2] The equivalent sums of money in euros per month were also printed in each card, so that respondents could express their answers in either euros per month or euros per day. For each bid, respondents had to state whether: (1) 'It would be definitely high enough', (2) 'It would be definitely not high enough', or (3) 'I am not sure whether it would be high enough or not'. Then, in a follow-up question, those respondents who stated that €15 would be definitely not high enough were directly asked to specify the minimum amount of money they would require to provide one additional hour of care per day. Conversely, those respondents who provided a zero answer (i.e. they stated that €0 would be definitely high enough) were subsequently asked for the reason why they needed no monetary compensation to provide an extra hour of care per day.

In the sample of carers, the three $WTA_{hypothetical}$ questions were followed by a variety of care-related questions. First, they were asked how many people they took care of. Those carers who took care of more than one person, were told to take into account exclusively the main care recipient (i.e. the person they looked after most of their time) when answering the subsequent questions. Next, they were asked about the relationship with the care recipient and if they had thought of that person when they answered the three $WTA_{hypothetical}$ questions. The purpose of this latter question was to ascertain whether carers were able to abstract from their own situation when they had to deal with the hypothetical caring scenario. Other questions of this section asked for the care recipient's age, the number and type(s) of care recipient's disability(ies), whether the carer and the care recipient lived in the same house, whether the carer provided

[2] This range of values was chosen because the pilot study revealed that it was plausible (e.g. it did not lead to a large rate of protest zeros and the maximum WTA value was €12).

informal care without receiving assistance from other carers, the duration and frequency of care, the number of hours of care per day (as a whole and also depending on the type of task), and the main reason why the carer provided informal care. Moreover, the carers also had to rank the different types of informal care activities they used to provide (from the least preferred task to the most preferred task), in a similar way as they had previously ranked the activities described in the hypothetical scenario.

Additionally, carers' subjective burden (i.e. the self-perceived impact of caring on carers' quality of life) was assessed using the CarerQol-7D questionnaire (Brouwer *et al.*, 2006; Hoefman *et al.*, 2011). This instrument comprises seven dimensions: fulfillment, relational problems, mental problems, problems with combining daily activities, financial problems, support and physical problems. Each dimension can take three levels: (1) 'no', (2) 'some', (3) 'a lot of ', which makes possible to characterise a total of 2,187 informal caring situations. After completing the CarerQol-7D questionnaire, the carers were asked to rate on a vertical scale how satisfied they were with the informal care they provided. The endpoints of this scale (0-100) were labelled 'the lowest imaginable satisfaction with care' and 'the highest imaginable satisfaction with care', respectively.

After the questions related to the care provided, carers were asked to state the minimum compensation they would require if they had to spend an extra hour per day looking after their own care recipient (the full wording of this question is reported in Figure 3.2). In case there were several care recipients, this question was posed in relation to the main care recipient. For short, this valuation will be labeled 'WTA_{own}' in what follows. Obviously non-carers' questionnaires did not include the set of care-related questions. Instead, after the three $WTA_{hypothetical}$ questions, non-carers were asked whether they were informal carers in the past and whether they personally knew some informal carer.

Both types of questionnaires ended with a set of questions regarding demographic and socioeconomic factors (age, gender, nationality, place of residence, marital status, children younger than 18 years living at home, household size, education, employment status and income), religion, personality, health and life satisfaction. Three types of health-related questions were posed. First, participants were asked to rate their overall health as excellent, good, fair, poor or very poor. Second, respondents' health-related quality of life (HRQoL) was assessed by the SF-6D descriptive system. This measure provides a classification of health states through six domains: physical functioning, role limitation (which combines role-physical limitation and role-emotional limitation), social functioning, bodily pain, mental health and vitality. Each dimension can take several levels of severity (between 4 and 6, depending on the dimension), which allows to define up to 18,000 health states by selecting one level from each dimension. The SF-6D can be derived from either the SF-36 (Brazier *et al.*, 2002) or the SF-12 (Brazier and Roberts, 2004) health profiles. In this study we used the SF-6D version which comes from 11 items of the SF-36. In addition, respondents were asked whether they had been diagnosed with some chronic illness and, if so, to specify which illness(es) they suffered from. Finally, overall life satisfaction was measured on a scale with the same format as the satisfaction with care scale described above. In this case, the endpoints (0-100) were labelled 'the lowest imaginable satisfaction with life' and 'the highest imaginable satisfaction with life', respectively.

Figure 3.1. Description of the hypothetical scenario and wording of the $WTA_{general}$ question

Imagine that you have been caring for a loved person (e.g. parent, sibling, partner) for a year, who lives at home with you. Besides, you are the only carer who provides care to that person. Imagine that your care recipient:

- needs help to eat and drink.
- has both urinary and fecal incontinence and you have to change his/her nappies.
- is unable to perform most personal care activities (personal hygiene, getting dressed, etc.).
- needs assistance with mobility (at home and out of home) and to stand up.
- is unable to do most housework.
- has problems to live alone, needs help to manage money and drugs and to make some everyday decisions; puts up no resistance to the assistance that he/she receives from you.

Due to the health problems of that person, you have to look after him/her for 8 hours every day. More specifically, you spend that time carrying out the following activities:

- 3 hours performing personal care tasks: bathing, dressing, toileting, changing nappies, giving food and drink.
- 1 hour providing assistance with mobility: walking at home and at of home.
- 3 hours doing housework: cooking, cleaning, washing, ironing, sewing, gardening.
- 1 hour carrying out practical activities: administering medicines, handling money, telephoning, going to the doctor (when necessary), doing the shopping.

Suppose you should look after that person for one additional hour every day. Thus, you should have to spend 9 hours per day –instead of 8 hours- providing informal care. In addition, imagine that the Government would give you a monetary compensation for providing that extra hour of care. Given that situation, we ask you to think about the minimum sum of money that you would need to be compensated if you had to provide that additional hour of care every day. In order to help you answer this question, we are going to show you a series of cards with an amount of money printed on them. Please, sort these cards into three groups, depending on whether:

Figure 3.1 (continued)

1. This sum would be definitely high enough.
2. This sum would be definitely not high enough.
3. I am not sure whether this sum would be high enough or not.

We warn you to check your answers so as to avoid possible contradictions. For instance, you cannot state that €8 per day would be definitely high enough compensation and, at the same time, that €10 per day would not be definitely high enough or that you are not sure whether this sum would be high enough or not.

[Only for those who declared that €0 would be ‘definitely high enough’]
You have just stated that €0 would be ‘definitely high enough’ to be compensated for devoting one additional hour per day to take care of that person. Could you please tell us why you would not consider necessary to receive any monetary compensation for that extra hour of care per day?
1. Providing one additional hour of care per day would not involve so much effort so as to need to be compensated for it.
2. It is a matter of consciousness; I would feel pangs of remorse if I accepted any money for looking after a loved one.
3. Other reasons (please, specify): ………………………………………………………..

[Only for those who said that €15 per day would be ‘definitely not high enough’]
Please, could you indicate us what is the minimum daily (or montly) amount of money that you would consider necessary to be compensated for looking after that person for one extra hour every day?
…………€/day or ……….€/month

Figure 3.2. Wording of the WTA_{own} question

We ask you to consider all the aspects related to your own caring situation: your care recipient’ health status, the number of days you provide care, the number of hours of care per day, the caring tasks you usually carry out and the time you invest in them, the positive and negative effects that you experience as a result of being an informal carer, etc.). If you care for several care recipients, we ask you to think exclusively of the person you usually look after most of your time.
Now suppose your loved one would need you to care for him/her for one additional hour every day, and that the Government would give you a monetary compensation in exchange for providing that extra hour of care. If you were in that situation, what would be the minimum sum of money you would need to be compensated for that additional hour of care? In order to help you answer this question, we are going to show you a series of cards with an amount of money printed on them. Please, sort these cards into three groups, depending on whether:

Figure 3.2 (continued)

1. This sum would be definitely high enough.
2. This sum would be definitely not high enough.
3. I am not sure whether this sum would be high enough or not.

We warn you to check your answers so as to avoid possible contradictions. For instance, you cannot state that €8 per day would be definitely high enough compensation and, at the same time, that €10 per day would not be definitely high enough or that you are not sure whether this sum would be high enough or not.

[Only for those who declared that €0 would be 'definitely high enough']
You have just stated that €0 would be enough to be compensated for devoting one additional hour per day to take care of that person. Could you please tell us why you would not consider necessary to receive any monetary compensation for that extra hour of care per day?

1. Providing one additional hour of care per day would not involve so much effort so as to need to be compensated for it.
2. It is a matter of consciousness; I would feel pangs of remorse if I accepted any money for looking after a loved one.
3. Other reasons (please, specify): ………………………………………………………..

[Only for those who said that €15 per day would be 'definitely not high enough']
Please, could you indicate us what is the minimum daily (or montly) amount of money that you would consider necessary to be compensated for looking after that person for one extra hour every day?
…………€/day or ……….€/month

3.2.3. Hypotheses

We set out a number of hypotheses on the basis of the economic theory, previous research on carers' burden and on the application of the CV method for the valuation of informal care and health outcomes. The main hypotheses were formulated as follows:

Hypothesis 1. The rate of protest zeros would be higher in the sample of carers than in the sample of non-carers.

Hypothesis 2. Furthermore, carers would be more likely to state a protest zero in the WTA_{own} question than in any of the three $WTA_{hypothetical}$ questions. The reasoning behind this hypothesis is that ethical objections are more likely to arise when informal carers are asked about the person they look after than when they are asked about a fictitious person.

Hypothesis 3. Compared with non-carers, carers would demand less money in exchange for taking care of the hypothetical care recipient for one extra hour per day. By making this hypothesis we assumed that, as a result of the adaptation to their caring role, carers would regard the fact of providing one further hour of informal care as less unpleasant than non-carers would do. In consequence, carers would require a lower amount of money to be compensated for that additional hour of care.

Hypothesis 4. To accomplish logical consistency, in a strict sense, the ordinal ranking of the three $WTA_{hypothetical}$ values should satisfy the following condition (at least at the aggregate level): $WTA_{best} < WTA_{general} < WTA_{worst}$. Because those who stated the same value in two of the three $WTA_{hypothetical}$ questions or in the three of them cannot be regarded as inconsistent individuals, but rather as insensitive (or as indifferent) subjects , the prior criterion could be relaxed by considering that logical consistency was met if, on average, $WTA_{best} \leq WTA_{general} \leq WTA_{worst}$,. Therefore, WTA for the least preferred task and for the most preferred task should yield an upper and a lower bound for WTA, respectively. In other words, if WTA is consistent with respondents' preferences for different caring activities, the demanded compensation for one additional hour of care should be higher if that hour should be devoted to the most unpleasant task than if that time should be invested in the least unpleasant task.

Hypothesis 5. In addition, since non-carers might not have well-developed preferences for informal care (especially if they had no prior experience as informal carers), it might be expected that non-carers' valuations would be less reliable than carers' valuations, and that non-carers would be more likely to anchor in the cost of professional care when stating their WTA values. If so, the rates of logically inconsistent and insensitive respondents would be higher in the sample of non-carers than in the sample of carers.

Hypothesis 6. Given that the hypothetical scenario (which was the same for all respondents) and the caring situation of each carer are not comparable, a priori, it was unclear whether carers would report greater or lower values in the $WTA_{general}$ question than in the WTA_{own} question. However, it is interesting to examine the difference between the two valuations, since it might allow us to ascertain whether informal carers' preferences over the hypothetical situation were influenced (or biased) by their own experiences providing informal care, as suggested by results from preceding CV studies (Van den Berg *et al.*, 2008). In this regard, we compared the difference between the $WTA_{general}$ and WTA_{own} values in two groups of carers: those who bore in mind their own care recipient when they answered the $WTA_{hypothetical}$ questions and those who did not. The underlying assumption was that the difference between the two WTA values would be smaller in the former group than in the latter. As a result, the proportion of carers who demanded the same compensation in both the $WTA_{general}$ and WTA_{own} questions should be higher in the former group than in the latter.

A number of further hypotheses were established regarding the factors that we expected to be associated with the WTA values. To be exact:

Hypothesis 7. We hypothesised that, in the two samples, WTA would be positively associated with income, having a job and a high level of education. The main reasoning behind these assumptions is that providing one additional hour of informal care involves more opportunity costs for those individuals with a high socioeconomic level than for those who have a lower status (Van den Berg et al., 2005c). In any case, the expected positive association between income and WTA is more uncertain than it is between income and WTP because the latter is constrained by income (since it involves a payment) whereas the former is not (because it entails a compensation) (Brown and Gregory, 1999). The allegedly positive relationship between WTA and having a job is also explained by the fact that the provision of informal care is expected to entail more stress and burden when it has to be combined with a job (Molloy et al., 2008).

Hypothesis 8. Moreover, for carers, we expected WTA to be positively associated with those variables related to carers' burden (number of hours of care, being the only carer, caring for more than one person, caring for a blood relative or the partner, cohabiting with the care recipient, the number of care recipient's disabilities and looking after someone with a cognitive disability) (Montgomery et al., 1985; Pearlin et al., 1990; Hughes et al., 1999). Conversely, we hypothesised that WTA (at least WTA_{own}) would be negatively associated with the duration of care because we assumed that WTA would be lower in the most adapted carers and that the time providing informal care could be taken as a proxy for the adaptation to the caring role. We also expected to find a negative relationship between WTA and the CarerQol weighted score (to the extent that higher CarerQol weighted scores reflect better care-related quality of life) and between WTA and satisfaction with care.

Hypothesis 9. For non-carers we hypothesised that WTA would be lower if they had prior experience as informal carers and/or they knew some informal carer. Additionally, we also expected WTA to be negatively related to the SF-6D index because subjects with poor HRQoL are likely to experience a higher burden as a result of providing informal care (Brouwer *et al.*, 2006; Hastrup *et al.*, 2011).

3.2.4. Data analysis

Prior to all data analyses, each WTA measure was converted into a continuous variable by taking the midpoint between the lowest sum of money that each respondent stated that would be definitely high enough to be compensated and the highest amount that would be definitely not high enough compensation (or the respondent did not know whether it would be high enough or not). This transformation was not necessary when a respondent demanded more than €15 because, in this case, he/she was directly asked to specify the exact WTA value.

In case of zero WTA values, the reasons behind them were examined to distinguish between 'protest' zeros and 'true' zeros. Zero WTA values due to the following reason: 'Providing one additional hour of care per day would not involve so much effort so as to need to be compensated for it' were regarded as true zeros, because respondents who chose this option really attached a zero value to one extra hour of informal care. On the contrary, zero WTA values justified by the reason 'It is a matter of consciousness; I would feel pangs of remorse if I accepted any money for looking after a loved one' were considered to be protest zeros, since this argument denotes a rejection or a protest to the presented scenario due to ethical objections. When zero WTA values were justified by other reasons (which were detailed in an open-ended response option) it was assessed whether they were true or protest zeros. Respondents who stated a protest zero were excluded from the analyses, differentiating (in the sample of carers) between protest zeros provided in the $WTA_{hypothetical}$ questions and those stated in the WTA_{own} question, because the reasons that justify a zero WTA value under the hypothetical situation are not necessarily the same as the reasons that explain a protest answer under the own caring situation. Specifically, in the analyses involving the $WTA_{hypothetical}$ variables, those subjects who gave a protest zero in any of the three $WTA_{hypothetical}$ questions were excluded (regardless of whether or not they provided a protest zero in the WTA_{own} question). In the analyses concerning the WTA_{own} variable, those carers who gave a protest answer in the WTA_{own} question were left out (regardless of whether or not they stated a protest zero in any of the $WTA_{hypothetical}$ questions).

Different statistical analyses were performed to test the hypotheses listed in the previous subsection. Within-group differences were tested with the McNemar's test in case of proportions (hypothesis 2) and both the *t*-test and the Wilcoxon signed-rank test for continuous variables (hypothesis 4). Between-groups differences were tested with the Fisher's exact test for proportions (hypotheses 1, 5 and 6) and both the *t*-test and the Mann-Withney *U*-test for continuous variables (hypotheses 3 and 6).

To analyse the factors associated with WTA (hypotheses 7-9), an ordinary least squares (OLS) regression analysis was performed, with WTA in natural logarithm (ln) as dependent variable. This analysis was undertaken separately for each sample and for each WTA variable. The Ramsey test for omitted variables was performed and, to correct for heteroskedasticity, regressions were run with robust standard errors (RSE). All regressions comprised a set of demographic and socioeconomic variables as predictors of WTA (age, gender, marital status, children younger than 18 years living at home, educational attainment, employment status and monthly household income). Additionally, a set of care-related variables were entered into the regressions restricted to the sample of carers. These variables were: the number of hours of care per week, the duration of care (number of months providing informal care), the CarerQol weighted score[3] (which takes the value zero for the worst caring situation and the value 100 for the best situation), the number of care recipient's disabilities, and five categorical variables indicating whether the carer cared for more than one person, whether he/she received assistance from other carers, whether the carer looked after a blood relative or his/her partner, whether the carer and the care recipient cohabited and whether the care recipient had a mental disability. Moreover, the regression with WTA_{own} as dependent variable also incorporated a dichotomous variable indicating whether the carer thought of his/her actual care recipient when answering the three $WTA_{hypothetical}$ questions. Instead of the abovementioned set of care-related variables, the non-carers' WTA regressions included the SF-6D utility index[4] –which ranges from 0 ('death') to 1 ('perfect health')–, and two categorical variables indicating if the non-carers had prior experience as informal carers and if they knew some informal carer. So as to correct for non-linearity, age, the duration of care and the number of hours of care per week were entered into the corresponding regressions both without transformation and also in squared values divided by 100. For income, we accounted for the usual assumption of diminishing marginal utility of income by taking the natural logarithm (ln) of this variable (Van den Berg and Ferrer-i-Carbonell, 2007). An advantage of using the logarithmic transformation of income is that its coefficient stands for the income elasticity of WTA, since this latter variable was also transformed into a logarithmic scale. The selection of the regressors was based on previous research on the monetary valuation of informal care using the CV method (Van den Berg *et al.*, 2005b; Van den Berg *et al.*, 2005c; De Meijer *et al.*, 2010).

[3] The sum of the seven score domains of the CarerQol-7D results in a summary score (ranging from 7 to 21). The conversion of this unweighted score into a weighted score was made using the Dutch tariff for the CarerQol Dutch (Hoefman *et al.*, 2013), the only one that is available until now.

[4] To convert the health states characterized by the SF-6D descriptive system into a summary utility index a scoring algorithm needs to be used. In this study we used the Spanish tariff for the SF-6D (Abellán *et al.*, 2012).

3.3. RESULTS

3.3.1. Sample characteristics

The 202 carers and the 200 non-carers who participated in the study fully completed the questionnaire. Table 3.2 displays the characteristics of the two samples. In addition, for comparative purposes, the Spanish population of informal carers and the Spanish adult general population are also characterised.

Table 3.2. Respondent's and Spanish population's characteristics

Characteristic	Carers (n = 202)	Spanish informal carers[a]	Non-carers (n = 200)	Spanish general public[b]
General characteristics				
Mean (SD) age	47.8 (12.8)	52.9	42.3 (15.4)	46.1
Women (%)	68.3	83.6	50.0	50.9
Marital status (%)				
Married/living with a partner	68.3	76.2	57.0	56.7
Single	21.8	14.8	32.0	31.8
Divorced	6.4	2.1	5.5	4.1
Widowed	3.5	5.1	5.5	7.4
Child(ren) < 18 at home (%)	27.7	36.1	30.0	N/A
Educational attainment (%)				
Illiterate or low	45.1	60.1	38.5	30.1
Medium	28.2	10.5	26.0	45.1
High	26.8	7.0	35.5	24.7
Employment status (%)				
Employed/self-employed	45.5	26.0	55.0	43.7
Unemployed	20.8	7.3	19.5	16.5
Homemaker	22.8	44.2	9.0	25.8
Student	2.0	1.4	5.5	
Oher situations	8.9	21.1	11.0	14.0
Monthly household income (%)				
< €900	10.9	N/A	11.5	23.5
€901-1,500	33.2	N/A	31.0	28.0
€1,501-2,000	31.2	N/A	30.0	19.4
> €2,000	24.8	N/A	27.5	29.1
Mean (SD) SF-6D utility score	0.87 (0.13)	N/A	0.88 (0.12)	N/A
Mean (SD) life satisfaction score	73.6 (16.2)	N/A	73.3 (15.6)	N/A

Table 3.2 (continued)

Characteristic	Carers (n = 202)	Spanish informal carers[a]	Non-carers (n = 200)	Spanish general public[b]
Care-related characteristics				
Experience as informal carer (%)			36.5	N/A
Knew some carer (%)			74.5	N/A
Mean (SD) duration (years)	3.4 (1.1)	5.5		
Mean (SD) hours (per day)	4.5 (3.2)	10.6		
Mean (SD) frequency (days per week)	5.1 (2.3)	N/A		
No assistance from others (%)	11.9	47.2		
More than one care recipient (%)	10.9	15.4		
Care recipient is[c] (%)				
Parent	53.0	57.2		
Partner	6.9	16.8		
Son/daughter[d]	5.0	N/A		
Sibling	5.9	3.0		
Grandparent	16.8	5.5		
Other[d]	12.4	16.6		
Cohabit (%)	47.0	60.7		
Number of care recipient's disabilities (%)				
1	5.5	N/A		
2	6.4	N/A		
≥ 3	88.1	N/A		
Type of disability[e] (%)				
Physical	94.6	N/A		
Cognitive	62.9	N/A		
Physical and cognitive	57.9	N/A		
Mean (SD) CarerQol weighted score	85.2 (13.8)	N/A		
Mean (SD) satisfaction with care score	77.8 (18.2)	N/A		

N/A: Not available. SD: Standard deviation.

[a] The characteristics of the Spanish informal carers come from the Survey on Informal Care 2004 (IMSERSO, 2004).

[b] The characteristics of the Spanish general population were obtained from the Living Conditions Survey 2009 (National Statistics Institute, 2009).

[c] In those cases in which the carer was in charge of more than one person, we considered the relationship with the main care recipient.

[d] The statistics for the Spanish population of informal carers do not detail the proportion of informal carers who look after a son or a daughter. These carers (among others) are included in the category 'Other'.

[e] The category 'physical disabilities' refers to vision loss, hearing loss, mobility impairment, difficulties in carrying out personal care activities and difficulties in performing housework. The category 'cognitive disabilities' includes memory loss, learning disabilities, intellectual disabilities, communication disorders and relational problems.

The observed differences between the two samples in terms of age and gender are conditioned by the fact that a quota sampling procedure was used in the sample of non-carers but no quotas were established in the sample of non-carers. On the one hand, carers' mean age (47.8 years) was significantly higher than non-carers' mean age (42.3 years). The vast majority of carers and non-carers (88%) were younger than 65 years. On the other hand, while most carers (68.3%) were women, the sample of non-carers consisted of an equal proportion of men and women. Regarding marital status, the proportion of subjects who were married (or living with a partner) was higher among carers (68.3%) than among non-carers (57%), whereas the percentage of single individuals was lower in the former group (21.8%) than in the latter (32%). As far as the level of educational attainment is concerned, compared with non-carers, the sample of carers comprised a greater proportion of illiterate or less educated subjects (45.1% *vs*. 38.5%) and a lower proportion of highly educated subjects (26.8% *vs.* 35.5%). With respect to the employment status, the rate of rate of employed respondents was lower in the sample of carers (45.5%) than in the sample of non-carers (55%), whereas the proportion of homemakers was much higher in the former group (22.8%) than in the latter (9%). Approximately 20% of individuals in each group were unemployed. The distribution of monthly household income was similar in the two samples, with approximately 60% of subjects at the €901-2,000 threshold. Likewise, the mean SF-6D scores were almost identical in the two samples (0.87 and 0.88 in the samples of carers and non-carers, respectively), as well as the mean satisfaction with life scores (73 and 74 over 100, respectively). It is also worth noting that 36.5% of non-carers declared to have been informal carers in the past and that 74.5% of them knew of some informal carer.

The age and sex composition of the sample of non-carers resembled that of the Spanish adult general population, which was attributable to the age and sex quotas that were imposed in this sample. The main differences between our sample of non-carers and the Spanish adult general public were observed in terms of socioeconomic factors (i.e. education, employment and income). To be exact, in our sample of non-carers there were higher percentages of lower educated individuals than in the Spanish general public (38.5% *vs.* 30.1%, respectively), but, in turn, there was a higher proportion of highly educated respondents than in the Spanish general public (35.5% *vs.* 24.7%). With regard to the employment status, in our sample of non-carers there was a higher rate of employed subjects than in the Spanish general public (55% *vs.* 43.7%), but also a higher rate of unemployed people (19.5% *vs.* 16.5%). Moreover, compared with the Spanish general population, our sample of non-carers comprised a lower percentage of subjects whose monthly household income was lower than €900 (11.5% *vs.* 23.5%) and a higher percentage of subjects at the threshold €901-2,000 (61% *vs.* 47.4%).

Table 3.2 shows the main care-related characteristics of the sample of carers. On average, they had been providing informal care for 3.4 years and invested 4.5 hours per day in caring (with a mean frequency of 5 days per week). Most carers (88.1%) received assistance from other carers and only 10.9% of them looked after more than one care recipient. In general, when carers were in charge of more than one person, there were two care recipients. In those cases in which the carer looked after more than one person, the analysis was confined to the main care recipient. Approximately 50% of the carers cohabited with the care recipient, who was the carer's father or mother in more than half of the cases (53%). Besides, a considerable percentage of carers took care of a grandparent (16.8%). Most care recipients had various disabilities (indeed, 88.1% of them had three or more disabilities), being more frequently somatic than cognitive problems (almost 95% of care recipients had at least one type of physical disability, whereas around 63% of care recipients suffered from at least one type of cognitive disability). In addition, more than 50% of care recipients had simultaneously at least one type of each a physical and a cognitive disability. With regard to the measures of care-related quality of life, on average, both the mean CarerQol weighted score and the mean satisfaction with care score were quite high (85.2 and 77.8 on a 0-100 scale, respectively).

As can be seen in Table 3.2, our sample of carers was not representative of the Spanish population of informal carers, either in terms of demographic and socieconomic factors or in relation to care-related characteristics. In general, objective burden (as measured by the duration and the number of hours of care, as well as by the proportions of carers who did not receive assistance from other carers and who cohabited with the care recipient) were milder among our sample of carers than among the broad population of Spanish informal carers. Furthermore, on average, the carers in our sample were approximately 5 years younger than the Spanish informal carers as a whole, and the proportion of female carers was also lower in our sample. Large discrepancies were also observed regarding education, with a lower rate of illiterate and less educated carers and a substantially higher proportion of highly educated carers in our sample than in the broad population of Spanish informal carers.

Given that some care-related variables depend to a great extent on carers' age, the differences between two age groups were examined (Table 3.3). Carers' median age (i.e. 47 years) was taken as the cut-off point to divide the sample of carers into two balanced age groups. One group comprised those carers aged between 18 and 46 years (n = 99) and the other group consisted of those carers aged 47 years and above (n = 103). In general, the older group bore more objective burden than the younger group. Specifically, compared with carers aged 18-47 years, on average, older carers had been providing informal care for a longer time and they devoted more hours per day and more days per week to caring. In addition, the proportion of carers who received no assistance from other carers was significantly higher in the older group than in the younger one (around 17.5% *vs.* 6.1%, respectively). Furthermore, whilst carers who took care of a partner, parent or son/daughter more frequently belonged to the older age group, the proportion of those who looked after a grandparent was significantly higher in the younger group. However, the level of subjective burden was similar in the two age groups. In particular, the mean CarerQol weighted score and the mean satisfaction with care score were only slightly lower among older carers (83.9 and 76.6, respectively) than among the younger ones (86.6 and 79, respectively).

Table 3.3. Care-related characteristics according to age

	18-46 years (n = 99)	≥ 47 years (n = 103)	P-values[a]
Mean (SD) duration (years)	4.2 (2.7)	5.2 (2.5)	0.008
Mean (SD) hours (per day)	3.8 (2.9)	5.2 (3.3)	0.002
Mean (SD) frequency (days per week)	4.7 (2.4)	5.6 (2.1)	0.004
No assistance from others (%)	6.1	17.5	0.016
More than one care recipient (%)	10.1	11.7	0.823
Care recipient is (%)			
Parent	45.5	60.2	0.048
Partner	1.0	12.6	0.001
Son/daughter	3.0	6.8	0.332
Sibling	4.0	7.8	0.374
Grandparent	32.3	1.9	0.000
Other	14.1	10.7	0.524
Mean (SD) care recipient's disabilities	4 (0.22)	5 (0.19)	0.248
Type of disability (%)			
Physical	93.9	95.2	0.764
Cognitive	57.6	68.0	0.146
Physical and cognitive	52.5	63.1	0.154
Cohabit (%)	41.4	52.4	0.124
Mean (SD) CarerQol weighted score	86.6 (12.9)	83.9 (14.5)	0.176
Mean (SD) satisfaction with care score	79.0 (17.6)	76.6 (18.7)	0.340

SD: Standard deviation.
[a] Two-tailed *t*-test test for continuous variables; Fisher's exact test for categorical variables.

3.3.2. Analysis of protest zeros

Table 3.4 presents the frequencies and percentages of positive values, true zeros and protest zeros obtained in each WTA question. The overall incidence of zero WTA values was relatively low in the two samples (below 10% of responses except in the WTA_{own} question), as well as the frequencies of true zeros and protest zeros by separate. Protest zeros outnumbered true zeros in all WTA questions. Most 'protesting' respondents were considered as such because their zero WTA answers were justified by the reason 'It is a matter of consciousness; I would feel pangs of remorse if I accepted money in exchange for looking after a loved one'. Additionally, one carer who provided a zero value in the three $WTA_{hypothetical}$ questions and three carers who gave a zero value in the WTA_{own} question stated other reasons (in the open-ended response option). These reasons can be summarised as follows: 'I would not accept money in exchange for caring a loved one' and 'He/She (the care recipient) would do the same for me'. Given that both answers reflect a rejection or a protest to the presented scenario due to ethical objections, zero WTA responses justified by any of these two latter reasons were also regarded as protest zeros. A total of 15 (7.4%) carers and 13 (6.5%) non-carers provided a protest zero in at least one of the three $WTA_{hypothetical}$ questions. As previously explained, these respondents were dropped out from further analyses concerning the $WTA_{hypothetical}$ variables. In general, respondents who stated a protest zero in any $WTA_{hypothetical}$ question also gave the same kind of answer in all the other $WTA_{hypothetical}$ questions. The only exception was found with one carer who provided a protest zero in the $WTA_{general}$ question, while he stated a positive value in all the other WTA questions (including the WTA_{own} question). On the other hand, 22 (10.9%) carers stated a protest zero in the WTA_{own} question. Those carers were excluded from the subsequent analyses involving the WTA_{own} variable, regardless of whether or not they provided the same type of response in the $WTA_{hypothetical}$ questions. Amongt those 22 carers, 8 of them provided a positive value in the three $WTA_{hypothetical}$ questions.

The proportions of protest zeros in the three $WTA_{hypothetical}$ questions were higher in the sample of carers (approximately 7%) than in the sample of non-carers (between 5.5% and 6.5%, depending on the question), but the differences between the two samples were not significant (Table 3.5), thus rejecting hypothesis 1. On the contrary, hypothesis 2 was confirmed because, within the sample of carers, the incidence of protest zeros was significantly higher in the WTA_{own} question than in any of the three $WTA_{hypothetical}$ questions.

Table 3.4. Classification of responses

Carers	WTA_{best}	$WTA_{general}$	WTA_{worst}	WTA_{own}
Positive values	182 (90.1)	182 (90.1)	184 (91.1)	171 (84.7)
True zeros	6 (3.0)	5 (2.5)	4 (2.0)	9 (4.5)
Protest zeros	14 (6.9)	15 (7.4)	14 (6.9)	22 (10.9)

Non-carers	WTA_{best}	$WTA_{general}$	WTA_{worst}
Positive values	185 (92.5)	185 (92.5)	187 (93.5)
True zeros	3 (1.5)	2 (2.1)	2 (1.0)
Protest zeros	12 (6.0)	13 (6.5)	11 (5.5)

The figures are frequencies (percentages).

Table 3.5. Tests of differences related to the protest zeros

Comparisons	Measures	*P*-values[a]
Between-groups	$WTA_{general}$	0.845
	WTA_{worst}	0.680
	WTA_{best}	0.840
Within-group (Carers)	WTA_{own} *vs.* $WTA_{general}$	0.039
	WTA_{own} *vs.* WTA_{worst}	0.008
	WTA_{own} *vs.* WTA_{best}	0.008

[a] Fisher's exact test for between-groups comparisons; McNemar's test for within-group comparisons.

3.3.3. WTA values in reference to the hypothetical scenario

Table 3.6 displays the main descriptive statistics for each $WTA_{hypothetical}$ variable. As can be seen, within each group, the comparison among the mean/median values of the three variables confirmed the expected ranking (i.e. $WTA_{best} < WTA_{general} < WTA_{worst}$). This means that, at the aggregate level, both carers' and non-carers' WTA responses satisfied the criterion of logical consistency. Carers' mean WTA values (€5.3, €6.4 and €7.5 for WTA_{best}, $WTA_{general}$ and WTA_{worst}, respectively) were only marginaly lower (and, therefore, no significantly different) than non-carers' mean WTA values (€5.6, €6.5 and €7.9, respectively). Median WTA_{best} and WTA_{worst} values were also lower among carers (€4.5 and €7, respectively) than among non-carers (€5.5 and €9, respectively), while equal median $WTA_{general}$ values were obtained in the two groups (€5.5). The maximum WTA value was €20 in all cases except for the maximum WTA_{best} value elicited in the sample of carers, which was lower (€13.5). The distributions of the three $WTA_{hypothetical}$ variables (as well as the distribution of the WTA_{own} variable) are graphically presented in Appendix 3B.

Consistently with these results, the Mann-Withney *U*-test revealed no significant between-groups differences for any of the three $WTA_{hypothetical}$ variables (Table 3.7). This rejects hypothesis 3 (that non-carers' $WTA_{hypothetical}$ values would be higher than carers' $WTA_{hypothetical}$ values). However, significant within-group differences were found between whatever two of the three $WTA_{hypothetical}$ variables (i.e. $WTA_{general}$ *vs.* WTA_{worst}, $WTA_{general}$ *vs.* WTA_{best} and WTA_{worst} *vs.* WTA_{best}; p = 000 in all cases), confirming the hypothesis of logical consistency (hypothesis 4), at the aggregate level.

Table 3.6. WTA values, descriptive statistics (€/hour/day)

Carers	WTA_{best}	$WTA_{general}$	WTA_{worst}
Mean	5.3	6.4	7.5
SD	(2.6)	(3.1)	(3.0)
95% CI	(4.9, 5.7)	(5.9, 6.8)	(7.1, 8.0)
Median	4.5	5.5	7.0
IQR	(3.5, 7.0)	(4.5, 9.0)	(5.5, 9.0)
Maximum	13.5	20.0	20.0
N	187	187	187

Non-carers	WTA_{best}	$WTA_{general}$	WTA_{worst}
Mean	5.6	6.5	7.9
SD	(3.0)	(3.2)	(3.1)
95% CI	(5.1, 6.0)	(6.0, 7.0)	(7.5, 8.4)
Median	5.5	5.5	9.0
IQR	(3.5, 7.0)	(4.5, 9.0)	(5.5, 9.0)
Maximum	20.0	20.0	20.0
N	187	187	187

SD: Standard deviation. CI: Confidence interval. IQR: Interquartile range

Table 3.7. Tests of differences regarding the three $WTA_{hypothetical}$ variables

Comparisons	Measures	*P*-values	
		T-test	Non-parametric test[a]
Between-groups	$WTA_{general}$	0.718	0.657
	WTA_{worst}	0.263	0.205
	WTA_{best}	0.354	0.604

Table 3.7 (continued)

Comparisons	Measures	*P*-values	
		T-test	Non-parametric test[a]
Within-group (Carers)	$WTA_{general}$ *vs.* WTA_{worst}	0.000	0.000
	$WTA_{general}$ *vs.* WTA_{best}	0.000	0.000
	WTA_{worst} *vs.* WTA_{best}	0.000	0.000
Within-group (Non-carers)	$WTA_{general}$ *vs.* WTA_{worst}	0.000	0.000
	$WTA_{general}$ *vs.* WTA_{best}	0.000	0.000
	WTA_{worst} *vs.* WTA_{best}	0.000	0.000

[a] Mann-Withney *U*-test for between-groups comparisons; Wilcoxon signed-rank test for within-group comparisons.

Table 3.8 examines the logical consistency of the $WTA_{hypothetical}$ responses at the individual level. The number of inconsistent subjects was very low in the two groups. Indeed, no respondent reversed the expected ranking (i.e. nobody gave the ranking $WTA_{worst} < WTA_{general} < WTA_{best}$). Pairwise comparisons between two of the three $WTA_{hypothetical}$ variables show that 8 carers (4.3%) and 12 non-carers (6.4%) incurred in at least one inconsistency (meaning that either $WTA_{general} > WTA_{worst}$ or $WTA_{general} < WTA_{best}$ or $WTA_{worst} < WTA_{best}$). An overwhelming proportion of respondents stated the same value in at least two of the three $WTA_{hypothetical}$ questions. The proportion of subjects who equated WTA_{worst} with WTA_{best} (around 25% in each sample) was considerably lower than the percentage of respondents who equated $WTA_{general}$ with WTA_{worst} (approximately 40%), on the one hand, and $WTA_{general}$ with WTA_{best} (nearly 50%), on the other hand. It is worth noting that almost a quarter of subjects in each group stated the same value in the three $WTA_{hypothetical}$ questions. Because of the large proportion of insensitive or indifferent respondents, the proportion of individuals who satisfied the ranking $WTA_{best} \leq WTA_{general} \leq WTA_{worst}$ (95.7% of carers and 93.6% of non-carers) was dramatically higher than the rate of strictly consistent respondents (i.e. those who accomplished the ranking $WTA_{best} < WTA_{general} < WTA_{worst}$) (28.9% of carers and 32.1% of non-carers). No significant differences were found between the two samples in terms of the proportions of consistent, inconsistent and insensitive subjects. Therefore, hypothesis 5 (that the rate of logically inconsistent and insensitive respondents would be higher among non-carers than among carers) was rejected.

Table 3.8. Logical consistency at the individual level

	Carers	Non-carers	Between-groups differences
$WTA_{general} < WTA_{worst}$	104 (55.6)	116 (62.0)	0.248
$WTA_{general} = WTA_{worst}$	79 (42.2)	67 (35.8)	0.244
$WTA_{general} > WTA_{worst}$	4 (2.1)	4 (2.1)	1.000
$WTA_{general} < WTA_{best}$	4 (2.1)	8 (4.3)	0.380
$WTA_{general} = WTA_{best}$	92 (49.2)	93 (49.7)	1.000
$WTA_{general} > WTA_{best}$	91 (48.7)	86 (46.0)	0.679
$WTA_{worst} < WTA_{best}$	1 (0.5)	1 (0.5)	1.000
$WTA_{worst} = WTA_{best}$	46 (24.6)	48 (25.5)	0.905
$WTA_{worst} > WTA_{best}$	140 (74.9)	138 (73.8)	0.906
$WTA_{best} < WTA_{general} < WTA_{worst}$	54 (28.9)	60 (32.1)	0.574
$WTA_{best} \leq WTA_{general} \leq WTA_{worst}$	179 (95.7)	175 (93.6)	0.492
$WTA_{worst} < WTA_{general} < WTA_{best}$	0 (0.0)	0 (0.0)	-
$WTA_{worst} \leq WTA_{general} \leq WTA_{best}$	46 (24.6)	45 (24.1)	0.902
All $WTA_{hypothetical}$ are equal	45 (24.1)	44 (23.5)	1.000
At least one inconsistency	8 (4.3)	12 (6.4)	0.491

The figures are frequencies (proportions), except for the last column (which are *p*-values from the Fisher's exact test). In both samples, the percentages were calculated by dividing the frequencies into 187 (the number of respondents in each samples who provided no protest zero in any of the three $WTA_{hypothetical}$ questions).

For further details, Tables 3.9 and 3.10 report the main descriptive statistics for WTA_{worst} and WTA_{best} by taking into account which specific types of activities were ranked as the least preferred task and as the most preferred task (under the hypothetical scenario), respectively. The vast majority of respondents (approximately 60% of carers and 70% of non-carers) regarded assistance with personal care as the most unpleasant task to be carried out in the hypothetical situation. Conversely, most carers and non-carers (around 66% and 68%, respectively) declared that the least unpleasant task would be practical activities. Mobility was ranked as the least preferred task by approximately 20% of carers and 10% of non-carers, and as the most preferred task by about 20% of carers and non-carers. Finally, housekeeping was considered to be the most unpleasant task by around 18% of carers and 17% of non-carers, and as the least unpleasant task by roughly 11% of subjects in each sample. Not surprisingly, in the sample of carers, there was a high correlation between the ranking of tasks described in the hypothetical scenario and the ranking of tasks performed in the own caring situation (p = 0.000). More specifically, for 114 (61%) carers the least preferred task coincided in the two situations, whereas for 113 (55.1%) carers the most preferred task was the same in both cases.

Since a same activity could be ranked as the most unpleasant task by some respondents and as the least unpleasant task by other subjects, two values were obtained for each task, depending on whether a given task was regarded as the least preferred task (Table 3.9) or as the most preferred task (Table 3.10). In general, the mean/median WTA values for each task were higher when the activity was considered to be the least preferred task than when the same task was ranked as the most preferred task. The only exception was found with the median WTA value for personal care in the sample of carers, which was lower when this activity was ranked as the most unpleasant task in the hypothetical situation (€7) than when it was considered to be the least unpleasant task (€8).

In the sample of carers, the mean/median WTA_{worst} values were comprised within the interval €7.3/€7 (for mobility) and €8/€9 (for practical activities), while the mean/median WTA_{best} values ranged from €5.2/€4.5 (for both mobility and practical activities) to €7.4/€8 (for personal care). In the sample of non-carers, the lowest ant the highest mean/median WTA_{worst} values were €7/€6.8 (for practical activities) and €9.4/€9 (for mobility), respectively, whereas the mean/median WTA_{best} values ranged from €5/€3.5 (for personal care) to €6.7/€5.5 (for housekeeping). On the other hand, the largest differences between the mean/median WTA_{worst} and WTA_{best} values for a same activity were obtained with the tasks practical activities (in the sample of carers) and personal care (in the sample of non-carers), followed by the task mobility (in both groups).

The between-groups comparisons in terms of WTA_{worst} and WTA_{best} depending on the least preferred task and the most preferred task, respectively, did not reveal significant differences, except when mobility was ranked as the least preferred activity (although in this case the difference was only significant at the 10% level). In particular, the mean/median WTA_{worst} values were lower in the sample of carers (€7.3/€7) than in the sample of non-carers (€9.4/€9). A sharp difference between the two groups was that, while mobility was the task with the lowest mean WTA_{worst} value in the sample of carers (€7.3), it was the activity with the highest mean WTA_{worst} value in the sample of non-carers (€9.4).

Table 3.9. WTA_{worst} values by types of tasks (€/hour/day)

Least preferred task		Personal care	Mobility	Housekeeping	Practical activities
Carers	Mean	7.6	7.3	7.4	8.0
	SD	(3.0)	(3.7)	(2.5)	(2.8)
	95% CI	(7.1, 8.2)	(6.1, 8.5)	(6.5, 8.3)	(5.5, 10.5)
	Median	7.0	7.0	7.0	9.0

Table 3.9 (continued)

Least preferred task		Personal care	Mobility	Housekeeping	Practical activities
Carers	IQR	(5.5, 9.0)	(4.5, 9.0)	(5.5, 9.0)	(4.5, 11.0)
	Maximum	20.0	13.5	13.5	11.0
	N (%)	108 (57.8)	38 (20.3)	34 (18.2)	7 (3.7)
Non-carers	Mean	7.9	9.4	7.1	7.0
	SD	(3.0)	(4.6)	(2.4)	(3.6)
	95% CI	(7.4, 8.4)	(7.2, 11.7)	(6.2, 7.9)	(1.3, 12.7)
	Median	9.0	9.0	7.0	6.8
	IQR	(5.5, 9.0)	(7.0, 13.5)	(6.3, 9.0)	(4.0, 10.0)
	Maximum	20.0	20.0	11.0	11.0
	N (%)	132 (70.6)	19 (10.2)	32 (17.1)	4 (2.1)
Between-groups differences (*p*-values)	*T*-test	0.489	0.063	0.578	0.614
	Mann-Withney *U*- test	0.328	0.076	0.947	0.560

SD: Standard deviation. CI: Confidence interval. IQR: Interquartile range.

Table 3.10. WTA_{best} values by types of tasks (€/hour/day)

Most preferred task		Personal care	Mobility	Housekeeping	Practical activities
Carers	Mean	7.4	5.2	5.5	5.2
	SD	(2.1)	(2.3)	(2.3)	(2.8)
	95% CI	(4.0, 10.8)	(4.5, 6.0)	(4.4, 6.6)	(4.7, 5.7)
	Median	8.0	4.5	5.5	4.5
	IQR	(5.8, 9.0)	(4.5, 7.0)	(4.5, 5.5)	(3.5, 7.0)
	Maximum	9.0	13.5	11.0	13.5
	N (%)	4 (2.1)	38 (20.3)	21 (11.2)	124 (66.3)
Non-carers	Mean	5.0	5.3	6.7	5.5
	SD	(3.5)	(3.2)	(3.5)	(2.8)
	95% CI	(-3.7, 13.7)	(4.3, 6.4)	(5.1, 8.4)	(5.0, 6.0)
	Median	3.5	4.5	5.5	5.5
	IQR	(2.5, 9.0)	(2.5, 7.0)	(4.5, 9.0)	(3.5, 7.0)
	Maximum	9.0	18.0	13.5	20.0
	N (%)	3 (1.6)	37 (19.8)	20 (10.7)	127 (67.9)
Between-groups differences (*p*-values)	*T*-test	0.312	0.875	0.196	0.479
	Mann-Withney *U*- test	0.271	0.598	0.376	0.361

SD: Standard deviation. CI: Confidence interval. IQR: Interquartile range.

3.3.4. Comparison between carers' WTA values under the hypothetical and the actual caring situations

This subsection compares carers' $WTA_{general}$ with WTA_{own} values. Common descriptive statistics for both variables are shown in Table 3.11. The information is given for the whole sample of carers and also for two groups: those who said that they thought of their own care recipient when they answered the three $WTA_{hypothetical}$ questions and those who did not. First considering the whole sample of carers, the distributions of $WTA_{general}$ and WTA_{own} were found to be significantly different. More specifically, the mean/median WTA_{own} values (€5.2/€4.5) were lower than the mean/median $WTA_{general}$ values (€6.5/€5.5) (p = 000 with both the t-test and the Wilcoxon signed-rank test). The mean/median difference between the $WTA_{general}$ and WTA_{own} values was €1.2/€0. The same overall profile (i.e. $WTA_{general}$ higher than WTA_{own}) was replicated in the two groups of carers that we considered. As expected, the difference between the two WTA values was smaller among those carers who bore in mind their loved one than among those who did not. Indeed, the median difference between the two values was €0 in the former group, while it amounted to €1 in the latter group. However, the difference between the $WTA_{general}$ and WTA_{own} values did not follow significantly different distributions in the two groups of carers, according to the Mann-Withney U-test (p = 0.194). Figure 3.3 displays, for both groups of carers, the distribution of the difference between the two WTA values.

Table 3.11. Comparison between carers' $WTA_{general}$ and WTA_{own} values (€/hour/day)

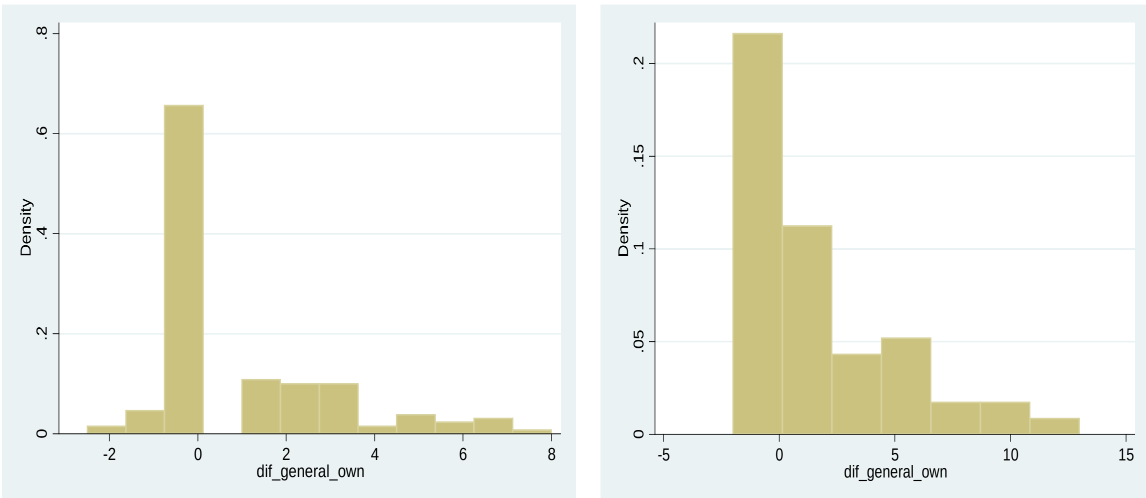
		All carers	The carer thought of his/her care recipient[a]	
			Yes	No
$WTA_{general}$	Mean	6.4	6.1	7.3
	SD	(3.1)	(2.9)	(3.5)
	95% CI	(5.9, 6.8)	(5.6, 6.5)	(6.3, 8.3)
	Median	5.5	5.5	5.5
	IQR	(4.5, 9.0)	(4.5, 7.0)	(4.5, 9.0)
	Maximum	20.0	13.5	20.0
	N (%)	187	137 (73.3)	50 (26.7)
WTA_{own}	Mean	5.2	5.0	5.7
	SD	(2.8)	(2.5)	(3.4)
	95% CI	(4.7, 5.6)	(4.5, 5.4)	(4.7, 6.8)
	Median	4.5	4.5	5.5
	IQR	(3.5, 7.0)	(3.5, 5.5)	(3.5, 7.0)
	Maximum	13.5	13.5	13.5
	N (%)	180	135 (75.0)	45 (25.0)

Table 3.11 (continued)

		All carers	The carer thought of his/her care recipient[a] Yes	No
$WTA_{general}$ *vs.* WTA_{own}	Mean (SD) difference	1.2 (2.1)	1.1 (1.8)	1.7 (2.8)
	Median (IQR) difference	0.0 (0.0, 2.0)	0 .0 (0.0, 2.0)	1.0 (0.0, 2.0)
	N (%)[b]	179	134 (74.9)	45 (25.1)
	Within-group difference (*p*-values)			
	T-test	0.000	0.000	0.000
	Wilcoxon signed-rank test	0.000	0.000	0.000
	Between-groups difference (*p*-values)			
	T-test		0.095	
	Mann-Withney *U* test		0.194	

SD: Standard deviation. CI: Confidence interval. IQR: Interquartile range.
[a] This division was made depending on whether the carers thought of their own care recipient when they answered the $WTA_{hypothetical}$ questions or not.
[b] The statistics provided for the comparison between $WTA_{general}$ and WTA_{own} are restricted to those carers who did not stated a protest zero in any of the two questions.

Figure 3.3. Difference between carers' $WTA_{general}$ and WTA_{own} values (€/hour/day)



Caregivers who thought of their own care recipient

Caregivers who did not think of their own care recipient

Based on individual-level data, Table 3.12 details the numbers and percentages of carers who stated lower and higher values in the $WTA_{general}$ question than in the WTA_{own} question, and of those who gave the same values in both situations. The information is offered for the whole sample of carers as well as for the same two groups that were distinguished in the previous table. A noteworthy finding is that more than half of carers (52%) demanded the same compensation for devoting one additional hour of care per day to look after their own care recipient than for taking care of the person described in the hypothetical scenario. It can be considered that those carers exhibited indifference between the hypothetical scenario and their actual caring situation. Not surprisingly, the proportion of carers who reported the same amount of money in the two questions was significantly higher among those who thought of their own care recipient when answering the $WTA_{hypothetical}$ questions (56%) than among those who abstracted from their actual situation (40%) (p = 0.084). Only a small fraction of carers (5.6%) stated a lower value in the $WTA_{general}$ question than in WTA_{own} question. Conversely, a considerable proportion of carers (42.4%) reported a higher value in the $WTA_{general}$ question than in the WTA_{own} question.

In summary, overall, hypothesis 6 was confirmed, in that, on average, the $WTA_{general}$ and WTA_{own} values were closer among those carers who thought of the person they looked after than among those who did not and, consequently, the proportion of carers who stated equal $WTA_{general}$ and WTA_{own} values.

Table 3.12. Comparison between carers' $WTA_{general}$ and WTA_{own} values at the individual level

		The carer thought of his/her care recipient	
	All carers	Yes	No
$WTA_{general} < WTA_{own}$	10 (5.6)	7 (5.2)	3 (6.7)
$WTA_{general} = WTA_{own}$	93 (52.0)	75 (56.0)	18 (40.0)
$WTA_{general} > WTA_{own}$	76 (42.4)	52 (38.8)	24 (53.3)
N (%)	179	134 (74.9)	45 (25.1)
Between-groups comparisons (*p*-values)			
$WTA_{general} < WTA_{own}$		0.714	
$WTA_{general} = WTA_{own}$		0.084	
$WTA_{general} > WTA_{own}$		0.116	

The values are frequencies (percentages). *P*-values come from the Fisher's exact test.

3.3.5. Regression analysis

To analyse the factors associated with the elicited WTA values, each WTA variable (in ln) was regressed against the set of explanatory variables listed in Table 3.13. Given that the ln of zero is not defined, the regression analyses were restricted to the positive WTA values. The resulting OLS estimates –coefficients and robust standard errors (RSE)– for the samples of carers and non-carers are displayed in Tables 3.14 and 3.15, respectively. The Ramsey test revealed no evidence of omitted variables for any model.

Table 3.13. List of explanatory variables used in the regression analysis

Variable name	Description
Only in regressions for carers' WTA	
Hours	Number of hours of informal care per week
Hours2	$Hours^2/100$
Duration	Number of months providing informal care
Duration2	$Duration^2/100$
No assist.	1 = the carer did not receive assistance from other carers; 0 = otherwise
Several	1 = the carer looked after more than one care recipient; 0 = otherwise
Family	1 = the carer and the care recipient were family related; 0 = otherwise
Cohabit	1 = the carer and the care recipient cohabited; 0 = otherwise
N. disab.	Number of care recipients' disabilities
Mental	1 = the care recipient had a cognitive disability; 0 = otherwise
CarerQol	CarerQol weighted score
Sat. care	Satisfaction with care score
Thought	1 = the carer thought of their own care recipient when answering the $WTA_{hypothetical}$ questions; 0 = otherwise
Only in regressions for non-carers' WTA	
Knew	1 = the non-carer knew some informal carer; 0 = otherwise
Exper.	1 = the non-carer had prior experience providing informal care; 0 = otherwise
SF6D	SF-6D utility index
In regressions for both carers' and non-carers' WTA	
Age	Respondent's age
Age2	$Age^2/100$
Female	1 = female respondent; 0 = male respondent
Married	1 = the respondent was married (or living together); 0 = otherwise
Children	1 = the respondent had children younger than 18 years living at home
Educ01	1 = the respondent was illiterate or had a low level of education; 0 = otherwise
Educ2	1= the respondent had a medium level of education; 0 = otherwise
Inc(ln)	Monthly household income (in ln)
Job	1 = the respondent had a paid job; 0 = otherwise

Table 3.14. OLS regressions for carers' WTA

	Ln(WTA_{best})		Ln($WTA_{general}$)		Ln(WTA_{worst})		Ln(WTA_{own})	
	Coeff.	RSE	Coeff.	RSE	Coeff.	RSE	Coeff.	RSE
Constant	1.601*	0.829	2.275***	0.764	1.813***	0.692	2.937***	0.845
Hours	-0.006	0.006	-0.003	0.006	0.005	0.005	-0.001	0.006
Hours2	0.011*	0.006	0.005	0.007	-0.004	0.005	0.005	0.006
Duration	0.157***	0.060	0.109**	0.054	0.039	0.047	0.086	0.060
Duration2	-1.169**	0.550	-0.797	0.513	-0.228	0.439	-0.645	0.562
No assist.	-0.239**	0.096	-0.189**	0.092	-0.158**	0.075	-0.294**	0.121
Several	-0.110	0.110	-0.124	0.121	-0.063	0.100	0.037	0.108
Family	0.011	0.103	-0.133	0.102	-0.086	0.094	0.081	0.074
Cohabit	0.026	0.084	0.012	0.089	0.002	0.072	-0.037	0.091
N. disab.	0.021	0.019	0.050**	0.021	0.037**	0.017	0.048**	0.020
Mental	-0.007	0.090	0.027	0.093	-0.009	0.081	-0.017	0.097
CarerQol	-0.008**	0.004	-0.008**	0.004	-0.008**	0.004	-0.009**	0.004
Sat. care	-0.000	0.002	0.000	0.002	0.003	0.002	0.001	0.002
Thought	-0.118	0.072	-0.151**	0.070	-0.060	0.064		
Age	0.040**	0.019	0.030*	0.017	0.031**	0.016	0.012	0.019
Age2	-0.045**	0.021	-0.036*	0.020	-0.037**	0.018	-0.012	0.021
Female	-0.057	0.075	0.002	0.078	-0.108*	0.060	-0.051	0.800
Married	-0.088	0.086	-0.099	0.087	-0.097	0.068	-0.189*	0.100
Children	-0.011	0.119	0.005	0.120	-0.058	0.115	0.025	0.128
Educ01	-0.192*	0.102	-0.141	0.097	-0.135*	0.078	-0.307***	0.107
Educ2	-0.248***	0.092	-0.210**	0.089	-0.191***	0.073	-0.238**	0.095
Inc(ln)	-0.031	0.093	-0.068	0.096	0.006	0.083	-0.138	0.105
Job	0.007	0.082	0.063	0.080	0.065	0.071	0.036	0.087
N	181		182		183		171	
Adj. R^2	0.153		0.151		0.115		0.165	
Ramsey test	0.464		0.404		0.162		0.330	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The figures shown for the Ramsey test are p-values.

Table 3.15. OLS regressions for non-carers' WTA

	Ln(WTA_{best})		Ln($WTA_{general}$)		Ln(WTA_{worst})	
	Coeff.	RSE	Coeff.	RSE	Coeff.	RSE
Constant	-0.385	0.885	-0.149	0.866	0.643	0.749
Age	-0.010	0.017	-0.023	0.018	-0.030**	0.014
Age2	0.012	0.018	0.024	0.018	0.032**	0.015
Female	-0.137*	0.080	-0.132	0.082	-0.074	0.060
Married	0.098	0.127	0.101	0.131	0.036	0.085
Children	-0.037	0.131	-0.011	0.135	0.008	0.087
Educ01	-0.095	0.105	-0.119	0.107	-0.040	0.081
Educ2	0.083	0.092	0.059	0.096	0.067	0.081

Table 3.15 (continued)

	Ln(WTA$_{best}$)		Ln(WTA$_{general}$)		Ln(WTA$_{worst}$)	
	Coeff.	RSE	Coeff.	RSE	Coeff.	RSE
Inc(ln)	0.359***	0.109	0.337***	0.107	0.280***	0.085
Job	-0.134	0.103	-0.143	0.112	-0.091	0.075
Knew	0.024	0.088	0.145	0.091	0.046	0.064
Exper.	0.109	0.091	0.199**	0.086	0.114*	0.068
SF6D	-0.489	0.416	-0.198	0.422	-0.111	0.361
N	185		185		185	
Adj. R^2	0.047		0.062		0.037	
Ramsey test	0.648		0.750		0.524	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The figures shown for the Ramsey test are p-values.

Focusing first on the estimates for the three carers' WTA$_{hypothetical}$ variables (Table 3.14), carers with higher care-related quality of life (as assessed by the CarerQol weighted score), those who received no assistance from other carers, as well as those with a medium level of education stated lower values in the three WTA$_{hypothetical}$ questions. On the other hand, carers' age and the three WTA$_{hypothetical}$ variables had an inverted *U*-shaped relationship, with WTA$_{general,}$ WTA$_{worst}$ and WTA$_{best}$ increasing up to approximately 47 years, 42 years and 44 years, respectively. Beyond those ages, the demanded compensation declined. Aside from these shared determinants, WTA$_{general}$ was found to have a positive association with the duration of care. The relationship between this latter variable and WTA$_{best}$ was quadratic, depicting an inverted *U*-shaped curve (instead of the expected *U*-shaped curve), although only for a short period of time –the maximum WTA$_{best}$ value was reached approximately at the seventh month of providing informal care. No significant association was found between WTA$_{worst}$ and the number of months providing informal care. None of the WTA$_{hypothetical}$ variables was significantly related to the number of hours of care, even when several specifications were tested. The required compensation in the WTA$_{general}$ and WTA$_{worst}$ questions increased with the number of care recipient's disabilities, but the type of disability (physical or cognitive) did not have a significant effect. Those carers who thought of their own care recipient when answering the three WTA$_{hypothetical}$ questions provided lower WTA values than those who did not (as expected), but this effect was significant only in the question that was posed in the first place (i.e. in the WTA$_{general}$ question). In addition, WTA$_{worst}$ was lower among female carers than among male carers, but this gender difference was significant at the 10% level.

The comparison between the variables associated with carers' $WTA_{general}$ and those variables related to WTA_{own} shows that both WTA values increased with the number of care recipient's disabilities. Furthermore, the two variables were negatively related to the CarerQol weighted score, to the fact of being the only carer and of having a medium level of education (as compared with having a high level of education). Apart from bein associated with these factors, WTA_{own} was significantly lower among married carers and among those with a low level of education. It is worth pointing out that neither the number of weekly hours of care nor the duration of care was significantly associated with WTA_{own}. Likewise, despite the significant relationship between age and the three $WTA_{hypothetical}$ values, WTA_{own} was not found to be significantly associated with age.

Turning to Table 3.15, the three $WTA_{hypothetical}$ values elicited in the sample of non-carers increased with income, contrary to what happened in the other sample. In particular, the following income elasticities were obtained: 0.34 for $WTA_{general}$, 0.28 for WTA_{worst} and 0.36 for WTA_{best} ($p < 0.01$ in the three cases). These figures mean that, if monthly household income rose by 1%, the required amount of money to be compensated for providing an extra hour of care per day would increase by 34%, 28% and 36%, respectively. Additionally, respondents who had been informal carers in the past provided higher $WTA_{general}$ and WTA_{worst} values than those who did not have that experience, which is the opposite of what we expected. Age had a *U*-shaped relationship with WTA_{worst}, with the latter variable decreasing up to approximately 47 years and increasing beyond that age. It should be remembered that the reverse direction effect was found for the three carers' $WTA_{hypothetical}$ values. Finally, comparing with men, women provided lower values in both the $WTA_{general}$ and WTA_{best} questions. No significant association was observed between non-carers' level of education and WTA.

In summary, among hypotheses 7-9, the regression analyses confirmed a number of them, although in some cases only partly. To be exact, the following hypothesis were accepted: that the income elasticity of WTA would be positive (this hypothesis was accepted only for non-carers); that WTA would be higher among respondents with a high level of education than among those with lower levels of education (it was only confirmed in the sample of carers); that those carers who thought of their own care recipient when they answered the $WTA_{hypothetical}$ questions would state lower WTA values than those who abstracted from their own caring situation (this effect was only significant for $WTA_{general}$); that WTA would increase with the number of care recipient's disabilities (this relationship was significant in all cases except for WTA_{best}) and that the relationship between WTA and the CarerQol weighted score would be negative (this association was significant in all cases). The remaining hypotheses were not confirmed, either because the signed was the opposite as expected (e.g. the effect of having been informal carer in the past, in the sample of non-carers), or because no significant relationship was found at all (e.g. the non-significant association between WTA and the number of hours).

3.4. DISCUSSION

Applications of stated preference methods (CV, CA and DCEs) for the valuation of informal care reported until now have focused on the carers' perspective (Van den Berg *et al.*, 2005c; Gustavsson *et al.*, 2010; Mentzakis *et al.*, 2010) and, in some cases, on both the carers' and the care recipients' points of view (Van den Berg *et al.*, 2005b; De Meijer *et al.*, 2010). There are several reasons why, aside from these two perspectives, it is also important to take account of the preferences of individuals who are neither informal carers nor care recipients (who have been called 'non-carers' throughout the paper). Specifically, as part of the general public, non-carers partly bear the indirect costs associated with informal care, they are potential (or former) informal carers and care recipients and, in theory, they may be a more objective (although supposedly less informed) source of valuation than informal carers and care recipients are (Van den Berg *et al.*, 2004). Along these lines, the main objectives of this study were to obtain a monetary value for informal care based on non-carers' stated preferences, as well as to compare this valuation with that elicited from a sample of actual informal carers.

The valuation approach used in this study was the CV method. To be exact, subjects in both samples were asked to state their minimum WTA (i.e. the minimum compensation they would require) for one additional hour per day looking after a fictitious care recipient. To test for logical consistency, three $WTA_{hypothetical}$ values were elicited: WTA for one extra hour per day of informal care in general terms ('$WTA_{general}$'), WTA for one extra hour per day undertaking the least preferred (or the most unpleasant) task ('WTA_{worst}'), and WTA for one additional hour per day performing the most preferred (or the least unpleasant) task ('WTA_{best}'). If the elicited WTA values were consistent with respondents' preferences for different caring activities, then the ordinal ranking of the three $WTA_{hypothetical}$ values should be $WTA_{best} < WTA_{general} < WTA_{worst}$, in a strict sense, or $WTA_{best} \leq WTA_{general} \leq WTA_{worst}$, allowing for equal values.

Overall, the results obtained in the sample of non-carers were quite acceptable and reflect that it is feasible to elicit a monetary value for informal care based on the stated preferences of this group (at least using WTA). On the one hand, all non-carers answered each WTA question (as well as the remaining questions) and the proportion of protest zeros was relative low (6.5%

(as well as the remaining questions) and the proportion of protest zeros was relative low (6.5% of non-carers gave a protest zero in at least one of the three $WTA_{hypothetical}$ questions and 5.5% of them provided this kind of response in the three cases). On the other hand, only a few non-carers (6.4%) incurred in at least one logical inconsistency. The least satisfactory result in this sample was that nearly a quarter of non-carers reported the same value in the three $WTA_{hypothetical}$ questions, which might suggest that non-carers' $WTA_{hypothetical}$ values did not properly reflect their preferences for different caring tasks. Because of the large proportion of insensitve respondents, only a third of non-carers met the strict criterion of logical consistency (i.e. $WTA_{best} < WTA_{general} < WTA_{worst}$).

A similar response profile was found in the sample of carers. Thus, the observed differences between the two groups were not statistically significant. More specifically, the mean WTA_{best}, $WTA_{general}$ and WTA_{worst} values were only marginally lower among carers (€5.3, €6.4 and €7.5, respectively) than among non-carers (€5.6, €6.5 and €7.9, respectively). The median WTA_{best} and WTA_{worst} values were also lower in the sample of carers (€4.5 and €7, respectively) than in the sample of non-carers (€5.5 and €9, respectively), whereas the same median $WTA_{general}$ value (€5.5) was obtained in the two groups.[5] Moreover, the rates of prostest zeros, logically consistent, inconsistent and insensitive respondents did not differ significantly between the two samples.

Van den Berg and Ferrer-i-Carbonell (2007) put forward asking CV questions for hypothetical care recipients so as to test whether the CV method is able to capture the subjective aspects of informal care, as the WBV method seems to do. This suggestion prompted us to compare carers' $WTA_{general}$ with the compensation they would require if they had to assist their own care recipient for one additional hour per day ('WTA_{own}'). This comparison revealed significant differences between the distributions of the two variables. To be exact, the mean/median WTA_{own} values (€5.2/€4.5)[6] were lower than carers' mean/median $WTA_{general}$ values (€6.4/€5.5). The distribution of WTA_{own} values roughly mirrored the distribution of carers' WTA_{best} values (mean/median: €5.3/€4.5). These results mean that, on average, carers required a lower monetary compensation for taking care of their own care recipient for one extra hour per day than if they had to spend one additional hour per day caring for the person described in the hypothetical scenario. This might suggest that the positive aspects associated

[5] Year 2012 values. The equivalent carers's mean/median WTA_{best}, $WTA_{general}$ and WTA_{worst} values in constant prices (year 2005 = 100) and purchasing power parity (PPP) (EU-27 = 100) are €4.7/€4, €5.7/€4.9 and €6.6/€6.2, respectively. The same correction lowers non-carers' mean/median WTA_{best}, $WTA_{general}$ and WTA_{worst} values to €5/€4.9, €5.7/€4.9 and €7/€8, respectively.
[6] The equivalent mean/median WTA_{own} values in constant prices and PPP are €4.6/€4.

with the provision of informal care might be weighted more heavily in the WTA_{own} question than in the $WTA_{general}$ question. This assumption is supported by the fact that carers were more prone to state a protest zero in the WTA_{own} question than in any of the three $WTA_{hypothetical}$ questions. This finding suggests that moral concerns were more likely to arise in the WTA_{own} question than in the $WTA_{hypothetical}$ questions. Aside form this reason, the rather small carers' WTA_{own} values could be explained by the relatively mild carers' burden.

Nevertheless, despite the significant differences between the distributions of carers' $WTA_{general}$ and WTA_{own} at the aggregate level, it should be remarked that over 50% of carers stated the same value in the $WTA_{general}$ question as in the WTA_{own} question. This is closely related to the fact that nearly 75% of carers thought of their own care recipient when they answered the $WTA_{hypothetical}$ questions (albeit they were warned to abstract from their own caring situation when answering these questions). Indeed, the proportion of carers who equated $WTA_{general}$ with WTA_{own} was significantly higher among those who bore in mind their own care recipient. Therefore, the large porportion of carers who provided the same value in the $WTA_{general}$ question as in the WTA_{own} question is explained, to a great extent, by the difficulty that most carers found to abstract from their own caring situation, even though all carers answered the $WTA_{general}$ question at the beginning of the survey.

This study also analysed, the sensitivity of WTA to respondents' personal characteristics and care-related factors. The most robust finding in the sample of non-caregvivers was that WTA increased with respondents' income, as expected. This relationship, however, was not significant in the sample of carers. The lack of a significant association between carers' WTA and income does not invalidate our estimates, insofar the hypothesised positive relationship between income and WTA is less evident than it is between income and WTP. This is explained because WTP is limited by budgetary constraints, while WTA is not (Brown and Gregory, 1999). There was also some evidence that WTA values were higher among those non-carers who were informal carers in the past than among those who did not have that experience, contrary to what was expected. As hypothesised, those carers with higher levels of care-related quality of life (as assessed by the CarerQol weighted score) reported lower WTA values. This negative relationship was found for all WTA variables (including WTA_{own}). Conversely, and contrary to our expectations, carers' WTA responses were insensitive to the number of hours of care they actually provided. This non-significant relationship between WTA and the number of hours of informal care has also been reported in previous CV studies (Van den Berg *et al.*, 2005c). Our findings imply that carers' WTA values were more sensitive to the subjective aspects of informal care (e.g. care-related quality of life) than to the objective aspects (e.g. the number of hours of care per week). In addition, the three carers' $WTA_{hypothetical}$ values (but not WTA_{own}) were found to have an inverted *U*-shaped association with carers' age. This could be related to the opportunity costs that the provision of informal care involves, which are higher for people of working age. Education was also a significant determinant of carers' WTA values. In particular, carers with a low or medium level of education reported lower $WTA_{hypothetical}$ and WTA_{own} values than those carers with a higher level of education. Again, this relationship could be explained by the opportunity costs associated with informal caring: given that these costs are more relevant for those carers with a high level of education (because they have more job opportunities), they are expected to require a higher compensation for devoting more time to provide informal care.

Compared with previous WTA research on the monetary valuation of informal care, the WTA values reported in this chapter are substantially lower. For instance, a CV study with a sample of Dutch informal carers obtained mean/median WTA values ranging from €9.5/€9.1 to €10.5/€9.1 per hour[7] (Van den Berg *et al.*, 2005b). The former values came from a sample of informal carers in charge of patients with rheumathoid arthritis, while the latter values were elicited from a heterogeneous sample of informal carers (i.e. carers who looked after patients with different health problems). Using the same data, the median WTP value also amounted to €9.1 per hour, but the mean WTP value was lower than the mean WTA value (€7.8 per hour in the rheumathoid artrhitis sample and €8.6 per hour in the heterogeneous sample). Similar values were found by Van den Berg and Ferrer-i-Carbonell (2007) using the WBV method (the mean values ranged from €8.8 to €10 per hour), despite this method is conceptually different from the CV approach. Notwithstanding, it should be noted that our WTA values are not directly comparable with those found in the above-mentioned studies because we used a different time frame. To be exact, in these CV studies, the WTA/WTP questions were posed on a weekly basis (i.e. in terms of an increase/reduction of informal care by one hour per week). Conversely, we used a daily basis (i.e. we asked for an increase by one hour per day). We opted for a daily basis instead of a weekly basis because the provision of one additional hour of informal care per week involves considerably less burden than the provision of one extra hour of informal care every day. For that reason, a priori, we expected our WTA values to be higher than those elicited in CV studies which have used a weekly time frame. However, our WTA values resulted to be lower than those reported in prior CV studies.

[7] The equivalent values in constant prices (year 2005 = 100) and PPP (EU-27=100) are a bit lower, which also applies to the subsequent monetary values cited in the Discussion. For example, €9.5 and €10.5 (in current Dutch prices of 2005) are equivalent to €9.1 and €10 (in constant prices and PPP), respectively.

The comparison with preceding studies on the monetary valuation of informal care using Spanish data is more limited because the vast majority of them have relied on revealed preference methods, especially on the replacement cost method (Oliva *et al.*, 2007; Oliva and Osuna, 2009). Indeed, we are only aware of one CV study (using WTP for a reduction of informal care by one hour per day) in which the monetary value of informal care was elicited from a sample of Spanish informal carers in charge of patients with Alzheimer's disease (Gustavsson *et al.*, 2010). In addition, informal carers from other three countries (United Kingdom, Sweden and the United States) also participated in this study. On average, Spanish informal carers were willing to pay £121.0 per hour a month (£4.0 per hour a day), which is equivalent to €4.7 per day (applying the average exchange rate for year 2010; £1.00 = €1.17) or to €4.1 in constant prices, PPP. It can be seen the proximity between this mean WTP value and our mean carers' WTA_{own} value €5.2 (€4.6 in constant prices, PPP). In any event, the WTP values reported by Gustavsson *et al.* (2010) are not directly comparable with ours, for two reasons. On the one hand, we did not restrict our study to informal carers of patients with a specific health condition such as Alzheimer's disease. On the other hand, WTA and WTP values cannot be directly compared because WTP involves a payment –and, thus, it is constrained by income-, whereas WTA entails a monetary compensation.

The fact that our WTA values were rather small rules out the possibility that respondents' WTA answers were anchored in the market price for professional care, because this is quite higher than respondents' WTA values. To be exact, the 2011 reference public price for the home assistance service in the Region of Murcia (the area where all respondents lived) was €14 per hour (Comunidad Autónoma de la Región de Murcia, 2011), which is equivalent to €12.4 in constant prices, PPP.

This research has a number of limitations, which should be viewed as potential paths for future research. First, the two samples had a relatively small size, were recruited in a nearby area and there might be a 'self-selection' bias, especially in the sample of carers. Furthermore, neither of the two samples was representative and, therefore, our results cannot be extrapolated. On the one hand, the sample of carers was not representative of the broad population of Spanish informal carers -on average, carers in our sample were younger and bore less objective burden. Nonetheless, it is worth noting that, unlike some other studies on the monetary valuation of informal care, which have restricted to informal carers in charge of patients with a particular health problem (e.g. Alzheimer's disease, rheumatoid arthritis, etc.), our sample of carers was more heterogeneous because it comprised informal carers who looked after care recipients with an array of disabilities. On the other hand, albeit the sample of non-carers was representative of the Spanish adult general public regarding age and sex, no representativeness was met with respect to other factors, such as income and educational attainment. More important, despite care recipients also belong to the general population, they were not represented in the sample of non-carers. This exclusion was done in an attempt to obtain more objective valuations than otherwise. Given that the care recipients were not represented in any of the two samples, it would be wrong to derive a public's WTA-based value of informal care by aggregating the carers' and the non-carers' WTA values elicited in this study. We think that it would be of great interest to obtain a monetary valuation for informal care from a sample of the general public (comprising informal carers, care recipients and other groups of the society).

A further reason why our results could have been influenced by a self-selection bias is that protest observations were excluded, as commonly done in CV studies. However, we are aware that this solution does not come without problems, because it may have similar consequences as if a part of the sample were left out non-randomly, and, therefore, the resulting reduced sample may be no longer representative of the original sample (Strazzera *et al.*, 2003). Furthermore, the fact of omitting protest answers can lead to biased and unefficient estimates (Atkinson *et al.*, 2012). Nonetheless, it must be taking into account that the rates of protest zeros obtained in our study were much lower than those reported in many CV studies, which are usually around 20-30% (Carson, 1991). Consequentely, the comparison between our WTA values before and after dismissing protesting observations casts small differences, with the former being only slightly lower than the latter.[8]

Another limitation is that, in the three $WTA_{hypothetical}$ questions, the same hypothetical scenario was presented to all respondents. Thus, the $WTA_{hypothetical}$ values reported in this study are contingent upon the selected scenario. Our proposal is to develop a multi-attribute system able to describe a variety of caring situations on the basis of the tasks to be provided, the time required to be invested in each task and the health state of the care recipient. This would enable us to estimate a set of monetary values for each possible situation. In addition, all WTA values were elicited for an increase in one hour of informal care per day and, therefore, no scope effects were tested. Future studies could examine the sensitivity of WTA responses according to the additional amount of time to be spent on caring and also to the time frame used (e.g. two extra hours per day, one extra hour per week).

[8] For instance, carers' and non-carers' mean $WTA_{general}$ values (€6.4 and €6.5, respectively) lower to €5.9 and €6.1, respectively, when protest WTA answers are counted, whereas mean WTA_{own} value (€5.2) lowers to €4.6. The remaining WTA values computed without dismissing protest zeros are provided upon request to the authors.

Finally, the same question order was kept in all questionnaires. This fact might lead to response order effects. Indeed, some of our results could be ascribable to this kind of bias. For instance, carers who thought of their own care recipient when answering the $WTA_{hypothetical}$ questions provided lower $WTA_{general}$ values, but this effect was not significant either in the WTA_{worst} or in the WTA_{best} questions. To avoid possible order effects, future studies should vary the question order, for example, by splitting the sample into several groups, assigning a different question order to each of them. It would be particularly interesting to investigate whether or not the fact of posing the WTA_{own} question before the $WTA_{hypothetical}$ questions leads to different results from those obtained when these three questions are formulated in the first place (as in the present case).

To conclude, this chapter has reported the first study for the monetary valuation of informal care based on the stated preferences of a sample of non-carers. In general, the results obtained were quite favourable (except for the large rate of insensitive subjects) and did not significantly differ from those obtained in the sample of carers, although carers' valuations seem to have been influenced by their own experience providing informal care.

3.5. APPENDIX 3A

The Questionnaires

PRESENTACIÓN

La presente encuesta forma parte de un estudio realizado por investigadores de la Universidad de Murcia pertenecientes al Grupo de Trabajo en Economía de la Salud (GTES), y está financiada con fondos públicos. El estudio tiene como finalidad estimar el valor monetario de los cuidados informales prestados a personas dependientes en España, lo cual ayudará a los poderes públicos en la toma de decisiones relativas a los dependientes y a sus cuidadores. Sus respuestas serán de gran ayuda para el éxito de este estudio, por lo que le agradecemos su colaboración. Le pedimos que responda de forma meditada y con total sinceridad a las preguntas que se le plantearán a lo largo del cuestionario. Puede tener la tranquilidad de que no existen respuestas correctas o incorrectas y la seguridad de que toda la información que nos proporcione será tratada confidencialmente y de manera anónima.

Las personas que cuidan de un familiar o conocido en situación de dependencia se denominan cuidadores informales o no profesionales. Este tipo de cuidadores son mayoritarios en España, donde se estima que en torno a un 80% de las personas dependientes recibe la ayuda de un familiar con quien convive.

A menudo, los cuidadores informales sufren efectos negativos sobre su salud y su calidad de vida. Entre otros, los cuidadores pueden experimentar los siguientes problemas: deterioro de su salud (dolores musculares, insomnio, sensación de cansancio, falta de fuerza física, depresión, necesidad de recibir tratamiento médico); dificultades para compaginar la labor de cuidador con otras responsabilidades (trabajo, cuidado de los hijos, tareas domésticas); conflictos con la pareja; reducción de contactos sociales y del tiempo dedicado a actividades de ocio; problemas de índole económica (no poder plantearse trabajar fuera de casa, renunciar a un empleo, necesidad de reducir la jornada laboral, dificultades financieras); etc. En cualquier caso, diversas investigaciones han demostrado que estos problemas se pueden ver compensados por una serie de aspectos positivos (satisfacción por el hecho de cuidar a un ser querido, sentirse útil e importante para esa persona, recibir afecto y aprecio por la ayuda prestada, etc.).

P0A. En la actualidad, ¿dedica Ud. de forma habitual al menos 1 hora a la semana a cuidar de alguna persona allegada (familiar, amigo/a o vecino/a) que, debido a su estado de salud, necesite ayuda para realizar alguna de las siguientes tareas de la vida diaria?
[*Mostrar tarjeta 1*]

1. ☐ Sí [*PASAR A **P0B***]
2. ☐ No [*ENTREGAR UN CUESTIONARIO PARA **NO CUIDADORES***]

P0B. ¿Podría Ud. decirme cuántas horas a la semana dedica habitualmente a cuidar de esa persona (o a esas personas, en caso de que Ud. cuide a más de una)?

1. ☐ Menos de 10 horas a la semana
2. ☐ Entre 10 horas y 20 horas a la semana
3. ☐ Más de 20 horas a la semana

[*ENTREGAR UN CUESTIONARIO PARA **CUIDADORES** Y REGISTRAR EN EL ESTADILLO LA CUOTA A LA QUE SE ASIGNA EL ENCUESTADO*]

CUESTIONARIO PARA CUIDADORES

VALORACIÓN DE LOS CUIDADOS INFORMALES (I)

En primer lugar, le vamos a plantear una serie de situaciones hipotéticas, es decir, situaciones que no se corresponden necesariamente con su situación actual. Por tanto, le pedimos que haga un esfuerzo y que, al responder a estas preguntas, tenga siempre presente la situación que le vamos a describir en cada caso, tratando de olvidarse de su situación personal.

P1A. Imagine que, desde hace un año, Ud. cuida de un ser querido próximo (p.ej. su padre o madre, hermano o hermana, cónyuge o pareja), que vive en el mismo domicilio que Ud. En concreto, Ud. dedica a esta labor 8 horas al día, y, además, es la única persona que cuida de su ser querido. Suponga que el estado de salud de esa persona es el que se describe en la tarjeta 2 [*Mostrar tarjeta 2*]. A la vista de las necesidades de esa persona, imagine que diariamente Ud. emplea esas 8 horas del modo que se indica en la tarjeta 3: [*Presentar la tarjeta 3*].
Imagine que esa persona necesitara que Ud. cuidara de ella durante una hora más cada día. Por tanto, ahora debería dedicar 9 horas diarias a su cuidado, en lugar de 8 horas. Suponga también que el Gobierno le daría una compensación económica por prestar esa hora extra. Ante esta situación, le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por tener que cuidar de esa persona durante una hora adicional diaria. Para ayudarle a responder, a continuación le presentaremos varias tarjetas con una cantidad de dinero impresa en ellas. Por favor, le pedimos que forme tres grupos con estas tarjetas, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P1B. Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P1C. En cualquier otro caso, saltar a P2*].

P1B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]

Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por cuidar de esa persona durante una hora más al día. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Cuidar de esa persona durante una hora más al día no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello.
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido).
3. ☐ Otros motivos (especificar): ………………………………………

P1C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al día*]

Por favor, ¿podría indicar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por cuidar de esa persona durante una hora más cada día?

...............€/día ó ………..€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

P2. A continuación le pedimos que considere los cuatro tipos de tareas que le volvemos a mostrar en la tarjeta 1 [*Enseñar tarjeta 1*] y que las ordene de más a menos incómoda de realizar, comenzando por aquella que Ud. crea que le resultaría más incómoda y terminando por la que Ud. crea que sería menos incómoda.

[*Para cada una de las tareas, marcar con una X en la columna correspondiente*]

	Más incómoda			Menos incómoda
	1	2	3	4
Cuidado personal	☐	☐	☐	☐
Movilidad	☐	☐	☐	☐
Tareas domésticas	☐	☐	☐	☐
Actividades prácticas	☐	☐	☐	☐

P3A. Ahora imagine de nuevo que se encuentra en la situación que le hemos planteado antes, según la cual Ud. lleva un año dedicando 8 horas diarias a cuidar a un ser querido, cuyo estado de salud es el mismo que le hemos descrito previamente [*Volver a mostrar la tarjeta 2*]. Encontrándose en esa situación, imagine que Ud. tuviera que prestar una hora más de cuidados al día (es decir, 9 horas diarias, en lugar de 8) y que, además, esa hora adicional diaria la tuviera que dedicar concretamente a … [*Especificar el tipo de tarea más incómoda, según la respuesta dada en P2*]. Suponga además que el Gobierno le diera una compensación económica por prestar esa hora extra. Le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por dedicar una hora más al día a realizar esa actividad**.** Para ayudarle a responder, le volvemos a mostrar una serie de tarjetas. Por favor, forme tres grupos, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P3B.Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P3C. En cualquier otro caso, saltar a P4A*].

P3B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]

Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por dedicar una hora más al día a realizar esa actividad. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Dedicar una hora más al día a esa tarea no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido)
3. ☐ Otros motivos (especificar): ……………………………………………….

P3C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al día*]

Por favor, ¿podría precisar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por pasar esa hora adicional diaria realizando……? [*Especificar la tarea más molesta, según la respuesta dada en P2*]

............€/día ó ……….€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

P4A. Suponga ahora que Ud. tuviera que prestar una hora más de cuidados al día (es decir, 9 horas diarias, en lugar de 8) y que, además, esa hora adicional diaria la tuviera que dedicar concretamente a … [*Especificar el tipo de tarea menos molesta, según la respuesta dada en P2*]. Suponga además que el Gobierno le diera una compensación económica por prestar esa hora extra. Le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por dedicar una hora más al día a realizar esa actividad**.** Para ayudarle a responder, le volvemos a mostrar una serie de tarjetas. Por favor, forme tres grupos, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P4B. Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P4C. En cualquier otro caso, saltar a P5*].

P4B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]
Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por dedicar una hora más al día a realizar esa actividad. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Dedicar una hora más al día a esa tarea no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido)
3. ☐ Otros motivos (especificar): ……………………………………………………….

P4C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al*

día]

Por favor, ¿podría precisar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por pasar esa hora adicional diaria realizando…? [*Especificar la tarea menos molesta, según la respuesta dada en P2*]

............€/día ó ……….€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

ASPECTOS RELACIONADOS CON LOS CUIDADOS

Las cuestiones anteriores estaban referidas a una situación hipotética. Ahora le pedimos que piense en su propia situación como cuidador/a y que responda a una serie de cuestiones al respecto.

P5. ¿A cuántas personas presta Ud. su ayuda, debido a algún problema de salud?

1. ☐ A un sola
2. ☐ A más de una (especificar a cuántas): ....

[*Si sólo cuida a una persona, pasar a P6. Si cuida a más de una persona, saltar a P8*]

P6. [*Si sólo cuida a una persona*]

¿Podría indicar la relación que tiene con esa persona a la que Ud. cuida?

1. ☐ Padre
2. ☐ Madre
3. ☐ Esposo/compañero sentimental
4. ☐ Esposa/compañera sentimental
5. ☐ Hijo
6. ☐ Hija
7. ☐ Hermano
8. ☐ Hermana

9. ☐ Abuelo
10. ☐ Abuela
11. ☐ Suegro
12. ☐ Suegra
13. ☐ Vecino
14. ☐ Vecina
15. ☐ Amigo
16. ☐ Amiga
17. ☐ Otra/s persona/s (especificar): ……………………………………………

P7. [Sólo *si cuida a una única persona*]

En algunas de las cuestiones anteriores le hemos preguntado acerca de la cantidad de dinero que sería suficiente para compensarle a cambio de pasar una hora diaria más cuidando a la persona cuyo estado de salud se ha descrito previamente. ¿Podría indicar si, a la hora de responder a dichas preguntas, ha pensado en la persona a quien Ud. cuida realmente?

1. ☐ Sí
2. ☐ No

[*Saltar a P11 en ambos casos*]

P8. [*Sólo si cuida a más de una persona. Respuesta múltiple; Señalar con una X todas las personas que mencione*]

¿Podría indicar la relación que tiene con cada una de esas personas a las que Ud. cuida?

1. ☐ Padre
2. ☐ Madre
3. ☐ Esposo/compañero sentimental
4. ☐ Esposa/compañera sentimental
5. ☐ Hijo
6. ☐ Hija
7. ☐ Hermano
8. ☐ Hermana
9. ☐ Abuelo
10. ☐ Abuela
11. ☐ Suegro
12. ☐ Suegra

13. ☐ Vecino
14. ☐ Vecina
15. ☐ Amigo
16. ☐ Amiga
17. ☐ Otra/s persona/s (especificar): ………………………………………………

P9. [*Sólo en caso de que cuide a más de una persona*]
Entre las distintas personas a quienes Ud. presta su ayuda, ¿normalmente, a cuál de ellas dedica un mayor número de horas a lo largo de la semana?

1. ☐ Padre
2. ☐ Madre
3. ☐ Esposo/compañero sentimental
4. ☐ Esposa/compañera sentimental
5. ☐ Hijo
6. ☐ Hija
7. ☐ Hermano
8. ☐ Hermana
9. ☐ Abuelo
10. ☐ Abuela
11. ☐ Suegro
12. ☐ Suegra
13. ☐ Vecino
14. ☐ Vecina
15. ☐ Amigo
16. ☐ Amiga
17. ☐ Otra/s persona/s (especificar): ………………………………………………

P10. [*Sólo si cuida a más de una persona*]
En algunas de las cuestiones anteriores le hemos preguntado acerca de la cantidad de dinero que sería suficiente para compensarle a cambio de pasar una hora diaria más cuidando a la persona cuyo estado de salud se ha descrito previamente. ¿Podría indicar si, a la hora de responder a dichas preguntas, Ud. ha pensado en alguna de las personas a las que cuida realmente?

1. ☐ Sí, y he pensado en la persona a quien cuido durante más horas a la semana
2. ☐ Sí, he pensado en las personas a las que cuido, pero en ninguna en particular
3. ☐ No he pensado en ninguna de las personas a quienes cuido

[***TRAS LA RESPUESTA A ESTA PREGUNTA, LEER LO SIGUIENTE: "POR FAVOR, LE PEDIMOS QUE EN LAS SUCESIVAS PREGUNTAS TENGA EN CUENTA EXCLUSIVAMENTE A LA PERSONA A LA QUE HABITUALMENTE CUIDA DURANTE UN MAYOR NÚMERO DE HORAS A LA SEMANA"***]

P11. ¿Podría especificar la edad de la persona a la que Ud. cuida?
[*Si no sabe indicar la edad exacta, preguntar si tiene 65 años o más o menos de 65 años*]

1. ☐ ☐ Años
2. ☐ No lo sé exactamente, pero tiene menos de 65 años
3. ☐ No lo sé exactamente, pero tiene 65 años o más

P12. ¿Vive Ud. en el mismo domicilio que esa persona?

1. ☐ Sí, de manera permanente
2. ☐ Por temporadas (fines de semana, cada cierto número de días o de semanas…)
3. ☐ No

P13. La ayuda que Ud. presta a esa persona …?

1. ☐ Es la única que recibe
2. ☐ No es la única, pero sí es la principal
3. ☐ Es una ayuda secundaria

P14. ¿Podría precisar cuánto tiempo lleva Ud. cuidando a esa persona?

1. ☐ Menos de un año
2. ☐ Entre un año y menos de dos años
3. ☐ Entre dos años y menos de cuatro años
4. ☐ Entre cuatro años y menos de ocho años
5. ☐ Ocho años o más

P15. ¿Con qué frecuencia presta Ud. su ayuda a esa persona?

1. ☐ Todos los días
2. ☐ Entre tres y seis días a la semana
3. ☐ Una o dos veces por semana
4. ☐ Una o dos veces al mes
5. ☐ Otra frecuencia (especificar): ………………………..

P16. Cada día que Ud. ayuda a esa persona, ¿cuántas horas le dedica, por término medio?

1. ☐ Menos de una hora
2. ☐ Entre una hora y menos de dos horas
3. ☐ Entre dos horas y menos de cinco horas
4. ☐ Entre cinco horas y menos de ocho horas
5. ☐ Ocho horas o más (especificar, si es posible): ………………………..

P17. Habitualmente, cada día que Ud. cuida a esa persona, ¿cuántas horas, aproximadamente, dedica a realizar cada uno de los cuatro tipos de tareas que se detallan en la siguiente tarjeta?
[*Mostrar tarjeta 1*]
Para ayudarle a responder, le pido que indique el intervalo en el que se encuentra el número de horas diarias que suele dedicar a cada una de esas tareas.

[*Para cada una de las tareas, marcar con una X en la columna correspondiente*]

	0 horas	Entre menos de 1 hora y menos de 2	Entre 2 horas y menos de 4	4 horas o más
Cuidado personal	☐	☐	☐	☐
Movilidad	☐	☐	☐	☐
Tareas domésticas	☐	☐	☐	☐
Actividades prácticas	☐	☐	☐	☐

P18. [*Sólo a quienes realicen más de una tarea de cuidados, según se desprenda de la pregunta anterior*]

A continuación, le pedimos que tenga en cuenta las distintas tareas que realiza habitualmente y que las ordene según lo incómodas o molestas que resulten para Ud., comenzando por la que considere más incómoda y finalizando por la menos incómoda.

[*Mostrar de nuevo la tarjeta 1 y marcar con una X en la columna correspondiente*]

	Más incómoda			Menos incómoda
	1	2	3	4
Cuidado personal	☐	☐	☐	☐
Movilidad	☐	☐	☐	☐
Tareas domésticas	☐	☐	☐	☐
Actividades prácticas	☐	☐	☐	☐

P19. Ahora le vamos a presentar una serie de afirmaciones relativas a los cuidados que Ud. presta. Por favor, lea estas afirmaciones con atención y, para cada una de ellas, marque con una X la opción que considere que mejor describe su actual situación como cuidador/a. En caso de que cuide a más de una persona, piense exclusivamente en los cuidados que presta a la persona a quien cuida durante un mayor número de horas semanalmente.

[*Los encuestados deben reflejar en el cuadernillo del encuestador sus respuestas a las preguntas P19A a P19G (marcando con una X la respuesta a cada una de las 7 afirmaciones). Entregar el cuadernillo para que respondan a estas preguntas, recogerlo y continuar con P20*].

P19A. Cuidar a esa persona es algo que me genera satisfacción

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P19B. Tengo problemas de relación con la persona a la que cuido (problemas de comunicación entre nosotros, él/ella se opone con agresividad física o verbal a la ayuda que le presto, etc.)

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P19C. Tengo problemas emocionales (estrés, miedo, pesimismo, depresión, bajo estado de ánimo, preocupación por el futuro, etc.)

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P19D. Tengo problemas para compaginar mi labor de cuidador/a con otras actividades diarias (tareas domésticas, trabajo, estudios, familia, actividades de ocio, etc.)

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P19E. Tengo problemas financieros como consecuencia de cuidar a esa persona

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P19F. Cuando lo necesito, recibo ayuda de otros (familia, amigos, vecinos, voluntarios, ONG´s, etc.) para cuidar a esa persona

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

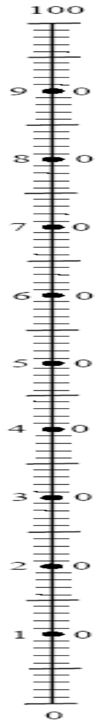
P19G. Tengo problemas de salud física (cansancio, mareos, somnolencia, dolor muscular, etc.) como consecuencia de cuidar a esa persona

1. ☐ No
2. ☐ Un poco
3. ☐ Mucho

P20. Los cuidadores pueden obtener cierta satisfacción o gratificación personal por el hecho de cuidar a una persona allegada. Nos gustaría conocer el grado de satisfacción global con los cuidados que Ud. presta. De nuevo, en caso de que Ud. cuide a más de una persona, le pedimos que piense exclusivamente en los cuidados que presta a la persona a quien cuida durante un mayor número de horas a la semana. Para ello, le vamos a presentar una escala de 0 a 100, donde el 100 se corresponde con el máximo grado de satisfacción imaginable que Ud. podría obtener con los cuidados y el 0 representa el mínimo grado de satisfacción imaginable. A la derecha de dicha escala aparece un recuadro; le pedimos que escriba dentro de él la puntuación (de 0 a 100) que Ud. daría a la satisfacción que obtiene con los cuidados que presta.

[*Los encuestados deben reflejar su respuesta en el cuadernillo del encuestador. Entregar el cuadernillo para que respondan a esta pregunta, recogerlo y continuar con P21*].

Máxima satisfacción
imaginable con los cuidados



Mínima satisfacción
imaginable con los cuidados

Su nivel de satisfacción global con los cuidados es:

P21. ¿Qué tipo de discapacidad/es tiene la persona a la que cuida Ud.? Si Ud. cuida a más de una persona, considere únicamente a la persona a quien cuida durante un mayor número de horas semanales.

[*Respuesta múltiple: Marcar todas las opciones que procedan*]

1. ☐ Problemas de visión
2. ☐ Dificultades para oír
3. ☐ Problemas de movilidad
4. ☐ Dificultades para el cuidado personal (asearse, vestirse…)
5. ☐ Dificultades para realizar las tareas del hogar
6. ☐ Problemas para relacionarse con los demás
7. ☐ Problemas de memoria (demencia, Alzheimer…)
8. ☐ Dificultades con actividades de aprendizaje básicas (leer, escribir, contar…)
9. ☐ Problemas de comunicación

10. ☐ Discapacidad intelectual
11. ☐ Otro tipo de discapacidad (especificar): ……………………………………………
12. ☐ Ninguna discapacidad

P22. ¿Cuál es el principal motivo por el que Ud. cuida a esa persona?
[*Respuesta múltiple: Marcar todas las opciones que correspondan*]

1. ☐ Lo considero una obligación moral
2. ☐ Esa persona no tiene ningún otro familiar o conocido que pueda cuidar de él/ella
3. ☐ Esa persona se opone a ser cuidado por alguien que no sea yo mismo/a
4. ☐ Económicamente, no puedo permitirme pagar a otra persona u organización para que preste los cuidados
5. ☐ Estoy cuidando de esa persona a la espera de que ésta sea admitida en un centro residencial
6. ☐ Otro motivo (especificar): ……………………………………………………………

VALORACIÓN DE LOS CUIDADOS INFORMALES (II)

P23A. A continuación le pedimos que considere todos los aspectos relacionados con los cuidados que Ud. presta a la persona que recibe sus cuidados (estado de salud de esa persona, número de horas/días de cuidados, tareas que presta y duración de las mismas, consecuencias negativas y positivas, que para Ud. tiene cuidar de él/ella, etc.). En caso de que Ud. cuide a varias personas, le volvemos a pedir que sólo tenga en cuenta los cuidados que presta a la persona a quien dedica un mayor número de horas a la semana. Ahora imagine que esa persona precisara de su ayuda durante una hora más cada día, y que el Gobierno le diera una compensación económica por prestar esa hora extra. Si Ud. se encontrara en esa situación, ¿qué cantidad mínima de dinero cree que le haría sentirse compensado/a por tener que prestar esa hora adicional diaria? Para ayudarle a contestar, le vamos a presentar varias tarjetas con una cantidad de dinero impresa en ellas. Por favor, forme tres grupos con estas tarjetas, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 €	0 €			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P23B. Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P23C. En cualquier otro caso, saltar a las preguntas de clasificación*].

P23B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]

Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por cuidar a esa persona durante una hora más al día. ¿Podría Ud. decirme por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Cuidar a esa persona durante una hora más al día no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a esa persona)
3. ☐ Otros motivos (especificar): …………………………………..

P23C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al día*]

Por favor, ¿podría indicar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por cuidar de esa persona durante una hora más cada día?

...............€/día ó ………..€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

PREGUNTAS DE CLASIFICACIÓN

Para finalizar con la encuesta, nos gustaría que respondiera a una serie de preguntas de carácter personal que servirán para poder interpretar mejor las respuestas que Ud. ha dado a lo largo del cuestionario.

C1. Sexo

1. ☐ Hombre
2. ☐ Mujer

C2. Edad: …..…..

C3. Nacionalidad: …………………………….

C4. Lugar de residencia: ………….…………….

C5. Estado civil

1. ☐ Soltero/a
2. ☐ Casado/a o pareja de hecho
3. ☐ Separado/a o divorciado/a
4. ☐ Viudo/a

C6A. ¿Tiene hijos/as menores de 18 años que convivan con Ud.?

1. ☐ Sí
2. ☐ No

C6B. ¿Cuántos/as?: …………

C7. En total, ¿por cuántas personas está formado su hogar (incluyéndose Ud. mismo/a)?: ………

C8. ¿Cuál es el mayor nivel de estudios terminados por Ud.?

1. ☐ Sin estudios
2. ☐ Primarios (EGB o similar)
3. ☐ Secundarios (BUP, COU, FP grado medio)
4. ☐ FP grado superior
5. ☐ Universitarios (diplomatura, licenciatura)

C9. ¿Cuál de las siguientes actividades describe mejor su situación actual?

1. ☐ Trabajando
2. ☐ En situación de desempleo
3. ☐ Percibiendo una pensión contributiva por jubilación o incapacidad permanente
4. ☐ Percibiendo otro tipo de pensión

5. ☐ Incapacitado/a para trabajar temporalmente
6. ☐ Estudiando
7. ☐ Dedicado/a principalmente a las labores del hogar
8. ☐ Realizando sin remuneración trabajos sociales o actividades benéficas
9. ☐ Otra situación

C10. Considere ahora los ingresos mensuales netos (es decir, descontando impuestos) de su hogar. ¿Podría señalar, de forma aproximada, en qué intervalo se encuentran éstos?

1. ☐ Menos de 900 €
2. ☐ Entre 901 y 1.500 €
3. ☐ Entre 1.501 y 2.000 €
4. ☐ Entre 2.001 y 3.000 €
5. ☐ Más de 3.000 €

C11. En relación a sus creencias religiosas, ¿podría decirme si Ud. se considera?

1. ☐ No creyente
2. ☐ Creyente, pero no practicante
3. ☐ Creyente y practicante

C12. ¿Cuál de los siguientes atributos considera que describe mejor su personalidad?

1. ☐ Pesimista
2. ☐ Realista
3. ☐ Optimista

C13. En comparación con otras personas de su misma edad, Ud. diría que su estado de salud actual es:

1. ☐ Excelente

2. ☐ Bueno
3. ☐ Regular
4. ☐ Malo
5. ☐ Muy malo

C14. A continuación le presentaremos seis preguntas relativas a su estado de salud actual. Por favor, léalas con atención y, para cada una de ellas, marque con una X la opción que considere que mejor describe su estado de salud en la actualidad.
Los encuestados deben reflejar en el cuadernillo del encuestador sus respuestas a las preguntas C14A a C14F (marcando con una X la respuesta a cada una de las 6 afirmaciones). Entregar el cuadernillo para que respondan a estas preguntas, recogerlo y continuar con C15].

C14A. En relación con su salud, señale la afirmación que mejor corresponde con su situación actual:

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 una hora)
4. ☐ Su salud le limita mucho para realizar esfuerzos moderados (p.ej. mover una mesa, pasar la aspiradora o caminar más de una 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

C14B. En relación con su salud física o emocional, señale la afirmación que mejor corresponde con su situación actual:

1. ☐ No tiene problemas con su trabajo u otras actividades cotidianas a causa de su salud física o de sus problema 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

C14C. En relación con su salud y sus actividades sociales (como visitar a amigos o familiares), señale la afirmación que mejor corresponde a su situación actual:

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

C14D. En relación con su salud, el dolor y el trabajo (fuera de casa o en las tareas del hogar), señale la afirmación que mejor corresponde a su situación actual:

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)

C14E. En relación con su estado de ánimo, señale la afirmación que mejor describe su situación actual:

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

C14F. En relación con su vitalidad, señale la afirmación que mejor describe su situación actual:

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

C15. ¿Podría decirnos si un médico le ha diagnosticado a Ud. alguna de las siguientes enfermedades crónicas?

[*Respuesta múltiple: Marcar todas las que procedan*]

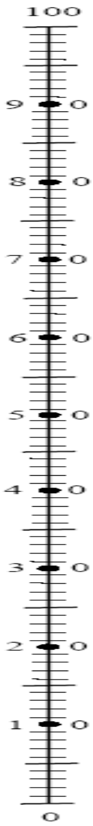
1. ☐ Artrosis/reumatismo
2. ☐ Alergia
3. ☐ Colesterol
4. ☐ Diabetes
5. ☐ Hipertensión arterial
6. ☐ Problemas del corazón/circulatorios
7. ☐ Problemas respiratorios (asma, bronquitis….)
8. ☐ Problemas de riñón, estómago o hígado
9. ☐ Dolor muscular (cuello, espalda…)
10. ☐ Osteoporosis
11. ☐ Depresión/problemas nerviosos

12. ☐ Tumor/cáncer
13. ☐ Enfermedades de la piel
14. ☐ Migraña/jaqueca
15. ☐ Otra/s enfermedad/es (especificar):
    ………………………………………………
16. ☐ Ninguna enfermedad

C16. Para finalizar con la encuesta, nos gustaría conocer cómo de satisfecho/a se siente Ud. considerando todos los aspectos de su vida en su conjunto (salud, familia, trabajo, situación financiera, amigos, ocio, etc.). Para ello, le vamos a presentar una escala de 0 a 100, donde el 100 equivale al máximo grado de satisfacción imaginable que Ud. podría sentir con todos los aspectos de su vida en su conjunto, y el 0 representa el mínimo grado de satisfacción imaginable. A la derecha de dicha escala aparece un recuadro; le pedimos que escriba dentro de él la puntuación de (0 a 100) que Ud. daría a su nivel de satisfacción global considerando todos los aspectos de su vida en su conjunto.

[*Los encuestados deben reflejar su respuesta en el cuadernillo del encuestador. Entregar el cuadernillo para que respondan a esta pregunta, recogerlo y continuar con C17*].

Máxima satisfacción
imaginable con su vida



Mínima satisfacción
imaginable con su vida

Su nivel de satisfacción global con su vida es:

C17. En una escala de 1 (muy fácil) a 10 (muy difícil), ¿podría indicar el grado de dificultad que ha encontrado a la hora de responder a este cuestionario?

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
1 2 3 4 5 6 7 8 9 10

Comentarios/Sugerencias:

………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………

MUCHAS GRACIAS POR SU COLABORACIÓN

CUESTIONARIO PARA NO CUIDADORES

VALORACIÓN DE LOS CUIDADOS INFORMALES (I)

En primer lugar, le vamos a plantear una serie de situaciones hipotéticas, es decir, situaciones que no se corresponden necesariamente con su situación actual. Por tanto, le pedimos que haga un esfuerzo y que, al responder a estas preguntas, tenga siempre presente la situación que le vamos a describir en cada caso, tratando de olvidarse de su situación personal.

P1A. Imagine que, desde hace un año, Ud. cuida de un ser querido próximo (p.ej. su padre o madre, hermano o hermana, cónyuge o pareja), que vive en el mismo domicilio que Ud. En concreto, Ud. dedica a esta labor 8 horas al día, y, además, es la única persona que cuida de su ser querido. Suponga que el estado de salud de esa persona es el que se describe en la tarjeta 2 [*Mostrar tarjeta 2*]. A la vista de las necesidades de esa persona, imagine que diariamente Ud. emplea esas 8 horas del modo que se indica en la tarjeta 3: [*Presentar la tarjeta 3*].
Imagine que esa persona necesitara que Ud. cuidara de ella durante una hora más cada día. Por tanto, ahora debería dedicar 9 horas diarias a su cuidado, en lugar de 8 horas. Suponga también que el Gobierno le daría una compensación económica por prestar esa hora extra. Ante esta situación, le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por tener que cuidar de esa persona durante una hora adicional diaria. Para ayudarle a responder, a continuación le presentaremos varias tarjetas con una cantidad de dinero impresa en ellas. Por favor, le pedimos que forme tres grupos con estas tarjetas, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P1B. Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P1C. En cualquier otro caso, saltar a P2*].

P1B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]
Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por cuidar de esa persona durante una hora más al día. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Cuidar de esa persona durante una hora más al día no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello.
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido).
3. ☐ Otros motivos (especificar): ………………………………………

P1C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al*

día]

Por favor, ¿podría indicar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por cuidar de esa persona durante una hora más cada día?

...............€/día ó ………..€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

P2. A continuación le pedimos que considere los cuatro tipos de tareas que le volvemos a mostrar en la tarjeta 1 [*Enseñar tarjeta 1*] y que las ordene de más a menos incómoda de realizar, comenzando por aquella que Ud. crea que le resultaría más incómoda y terminando por la que Ud. crea que sería menos incómoda.

[*Para cada una de las tareas, marcar con una X en la columna correspondiente*]

	Más incómoda			Menos incómoda
	1	2	3	4
Cuidado personal	☐	☐	☐	☐
Movilidad	☐	☐	☐	☐
Tareas domésticas	☐	☐	☐	☐
Actividades prácticas	☐	☐	☐	☐

P3A. Ahora imagine de nuevo que se encuentra en la situación que le hemos planteado antes, según la cual Ud. lleva un año dedicando 8 horas diarias a cuidar a un ser querido, cuyo estado de salud es el mismo que le hemos descrito previamente [*Volver a mostrar la tarjeta 2*]. Encontrándose en esa situación, imagine que Ud. tuviera que prestar una hora más de cuidados al día (es decir, 9 horas diarias, en lugar de 8) y que, además, esa hora adicional diaria la tuviera que dedicar concretamente a … [*Especificar el tipo de tarea más incómoda, según la respuesta dada en P2*]. Suponga además que el Gobierno le diera una compensación económica por prestar esa hora extra. Le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por dedicar una hora más al día a realizar esa actividad**.** Para ayudarle a responder, le volvemos a mostrar una serie de tarjetas. Por favor, forme tres grupos, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P3B.Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P3C. En cualquier otro caso, saltar a P4A*].

P3B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]
Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por dedicar una hora más al día a realizar esa actividad. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Dedicar una hora más al día a esa tarea no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido)
3. ☐ Otros motivos (especificar): ………………………………………………….

P3C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al día*]

Por favor, ¿podría precisar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por pasar esa hora adicional diaria realizando……? [*Especificar la tarea más molesta, según la respuesta dada en P2*]

............€/día ó ……….€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

P4A. Suponga ahora que Ud. tuviera que prestar una hora más de cuidados al día (es decir, 9 horas diarias, en lugar de 8) y que, además, esa hora adicional diaria la tuviera que dedicar concretamente a … [*Especificar el tipo de tarea menos molesta, según la respuesta dada en P2*]. Suponga además que el Gobierno le diera una compensación económica por prestar esa hora extra. Le pedimos que piense cuál cree que sería la cantidad mínima de dinero que le haría sentirse compensado/a por dedicar una hora más al día a realizar esa actividad**.** Para ayudarle a responder, le volvemos a mostrar una serie de tarjetas. Por favor, forme tres grupos, según considere que el importe mostrado en ellas:

1. Seguro sería una compensación suficiente
2. Seguro sería una compensación insuficiente
3. No sabe si sería una compensación suficiente o no

Le rogamos que revise sus respuestas para evitar posibles contradicciones. Por ejemplo, no puede señalar que una compensación de 8 € diarios sería suficiente y al mismo tiempo indicar que 10 € al día serían insuficientes o no sabe si serían suficientes o no.

[*Mostrar las tarjetas aleatoriamente, tras haberlas barajado. Anotar las respuestas en la siguiente tabla, marcando con una X en la columna correspondiente*]

Importe diario	Importe mensual equivalente	Seguro sería suficiente	Seguro sería insuficiente	No sé si sería suficiente o no
0 € (*)	0 € (*)			
1 €	30 €			
2 €	60 €			
3 €	90 €			
4 €	120 €			
5 €	150 €			
6 €	180 €			
8 €	240 €			
10 €	300 €			
12 €	360 €			
15 € (*)	450 € (*)			

(*) [*Si responde que "Seguro sería suficiente" una compensación de 0 €, pasar a P4B. Si responde que "Seguro sería insuficiente" una compensación de 15 € diarios (450 € mensuales), formular la pregunta P4C. En cualquier otro caso, saltar a P5*].

P4B. [*Sólo a quienes digan que "Seguro sería suficiente" una compensación de 0 euros*]
Ud. acaba de decir que está seguro de que 0 euros serían suficientes para compensarle por dedicar una hora más al día a realizar esa actividad. ¿Podría Ud. decirnos por qué motivo no creería necesario recibir compensación económica alguna por esa hora adicional de cuidados?

1. ☐ Dedicar una hora más al día a esa tarea no me supondría una diferencia tan grande como para necesitar ser compensado/a por ello
2. ☐ Es una cuestión de conciencia (me sentiría mal aceptando dinero a cambio de cuidar a un ser querido)
3. ☐ Otros motivos (especificar): ……………………………………………….

P4C. [*Sólo a quienes respondan que "Seguro sería insuficiente" una compensación de 15 € al*

día]

Por favor, ¿podría precisar la cantidad mínima diaria (o mensual) que Ud. consideraría suficiente para ser compensado/a por pasar esa hora adicional diaria realizando…? [*Especificar la tarea menos molesta, según la respuesta dada en P2*]

............€/día ó ……….€/mes

[*Los encuestados pueden indicar el importe diario o mensual, según les resulte más sencillo*]

P5A. ¿Conoce Ud. a alguna persona allegada que, debido a su estado de salud, precise de la ayuda de otra/s persona/s para realizar algunas de las tareas de la vida diaria (cuidado personal, caminar, tareas domésticas, etc.)?

1. ☐ Sí
2. ☐ No

[*Si la respuesta es afirmativa, pasar a P5B. Si es negativa, saltar a P6*]

P5B. ¿Qué relación tiene con esa/s persona/s?

[*Respuesta múltiple: Marcar todas las opciones que correspondan*]

1. ☐ Padre	2. ☐ Madre
3. ☐ Esposo/compañero sentimental	4. ☐ Esposa/compañera sentimental
5. ☐ Hijo	6. ☐ Hija
7. ☐ Hermano	8. ☐ Hermana
9. ☐ Abuelo	10. ☐ Abuela
11. ☐ Suegro	12. ☐ Suegra
13. ☐ Vecino	14. ☐ Vecina
15. ☐ Amigo	16. ☐ Amiga

17. ☐ Otra/s persona/s (especificar): ……………………………………………

P5. ¿Ha pensado Ud. en esa persona (o en alguna de ellas, si conoce a varias) a la hora de responder a las cuestiones anteriores, en las que le he preguntado por la cantidad de dinero que sería suficiente para compensarle a cambio de pasar una hora diaria más cuidando a la persona cuyo estado de salud hemos descrito anteriormente?

1. ☐ Sí,
2. ☐ No

P6. Al inicio del cuestionario, Ud. ha dicho que actualmente no cuida personalmente a ningún familiar o conocido. Y en el pasado, ¿cuidó Ud. a algún familiar, amigo/a o vecino/a que, debido a su estado de salud, precisara de la ayuda de otra/s persona/s para realizar las actividades de la vida diaria?

1. ☐ Sí
2. ☐ No

PREGUNTAS DE CLASIFICACIÓN

Para finalizar con la encuesta, nos gustaría que respondiera a una serie de preguntas de carácter personal que servirán para poder interpretar mejor las respuestas que Ud. ha dado a lo largo del cuestionario.

C1. Sexo

1. ☐ Hombre
2. ☐ Mujer

C2. Edad: …..…..

C3. Nacionalidad: ……………………………..

C4. Lugar de residencia: ………….……………..

C5. Estado civil

1. ☐ Soltero/a
2. ☐ Casado/a o pareja de hecho
3. ☐ Separado/a o divorciado/a
4. ☐ Viudo/a

C6A. ¿Tiene hijos/as menores de 18 años que convivan con Ud.?

1. ☐ Sí
2. ☐ No

C6B. ¿Cuántos/as?: …………

C7. En total, ¿por cuántas personas está formado su hogar (incluyéndose Ud. mismo/a)?: ………

C8. ¿Cuál es el mayor nivel de estudios terminados por Ud.?

1. ☐ Sin estudios
2. ☐ Primarios (EGB o similar)
3. ☐ Secundarios (BUP, COU, FP grado medio)
4. ☐ FP grado superior
5. ☐ Universitarios (diplomatura, licenciatura)

C9. ¿Cuál de las siguientes actividades describe mejor su situación actual?

1. ☐ Trabajando
2. ☐ En situación de desempleo
3. ☐ Percibiendo una pensión contributiva por jubilación o incapacidad permanente

4. ☐ Percibiendo otro tipo de pensión
5. ☐ Incapacitado/a para trabajar temporalmente
6. ☐ Estudiando
7. ☐ Dedicado/a principalmente a las labores del hogar
8. ☐ Realizando sin remuneración trabajos sociales o actividades benéficas
9. ☐ Otra situación

C10. Considere ahora los ingresos mensuales netos (es decir, descontando impuestos) de su hogar. ¿Podría señalar, de forma aproximada, en qué intervalo se encuentran éstos?

1. ☐ Menos de 900 €
2. ☐ Entre 901 y 1.500 €
3. ☐ Entre 1.501 y 2.000 €
4. ☐ Entre 2.001 y 3.000 €
5. ☐ Más de 3.000 €

C11. En relación a sus creencias religiosas, ¿podría decirme si Ud. se considera?

1. ☐ No creyente
2. ☐ Creyente, pero no practicante
3. ☐ Creyente y practicante

C12. ¿Cuál de los siguientes atributos considera que describe mejor su personalidad?

1. ☐ Pesimista
2. ☐ Realista
3. ☐ Optimista

C13. En comparación con otras personas de su misma edad, Ud. diría que su estado de salud actual es:

1. ☐ Excelente
2. ☐ Bueno
3. ☐ Regular
4. ☐ Malo
5. ☐ Muy malo

C14. A continuación le presentaremos seis preguntas relativas a su estado de salud actual. Por favor, léalas con atención y, para cada una de ellas, marque con una X la opción que considere que mejor describe su estado de salud en la actualidad.
Los encuestados deben reflejar en el cuadernillo del encuestador sus respuestas a las preguntas C14A a C14F (marcando con una X la respuesta a cada una de las 6 afirmaciones). Entregar el cuadernillo para que respondan a estas preguntas, recogerlo y continuar con C15].

C14A. En relación con su salud, señale la afirmación que mejor corresponde con su situación actual:

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 una hora)
4. ☐ Su salud le limita mucho para realizar esfuerzos moderados (p.ej. mover una mesa, pasar la aspiradora o caminar más de una 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

---

C14B. En relación con su salud física o emocional, señale la afirmación que mejor corresponde con su situación actual:

1. ☐ No tiene problemas con su trabajo u otras actividades cotidianas a causa de su salud física o de sus problema 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

---

C14C. En relación con su salud y sus actividades sociales (como visitar a amigos o familiares), señale la afirmación que mejor corresponde a su situación actual:

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

---

C14D. En relación con su salud, el dolor y el trabajo (fuera de casa o en las tareas del hogar), señale la afirmación que mejor corresponde a su situación actual:

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)

C14E. En relación con su estado de ánimo, señale la afirmación que mejor describe su situación actual:

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

---

C14F. En relación con su vitalidad, señale la afirmación que mejor describe su situación actual:

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

C15. ¿Podría decirnos si un médico le ha diagnosticado a Ud. alguna de las siguientes enfermedades crónicas?

[*Respuesta múltiple: Marcar todas las que procedan*]

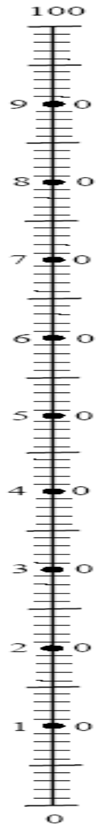
1. ☐ Artrosis/reumatismo
2. ☐ Alergia
3. ☐ Colesterol
4. ☐ Diabetes
5. ☐ Hipertensión arterial
6. ☐ Problemas del corazón/circulatorios
7. ☐ Problemas respiratorios (asma, bronquitis….)
8. ☐ Problemas de riñón, estómago o hígado
9. ☐ Dolor muscular (cuello, espalda…)
10. ☐ Osteoporosis
11. ☐ Depresión/problemas nerviosos

12. ☐ Tumor/cáncer
13. ☐ Enfermedades de la piel
14. ☐ Migraña/jaqueca
15. ☐ Otra/s enfermedad/es (especificar): ……………………………………………
16. ☐ Ninguna enfermedad

C16. Para finalizar con la encuesta, nos gustaría conocer cómo de satisfecho/a se siente Ud. considerando todos los aspectos de su vida en su conjunto (salud, familia, trabajo, situación financiera, amigos, ocio, etc.). Para ello, le vamos a presentar una escala de 0 a 100, donde el 100 equivale al máximo grado de satisfacción imaginable que Ud. podría sentir con todos los aspectos de su vida en su conjunto, y el 0 representa el mínimo grado de satisfacción imaginable. A la derecha de dicha escala aparece un recuadro; le pedimos que escriba dentro de él la puntuación de (0 a 100) que Ud. daría a su nivel de satisfacción global considerando todos los aspectos de su vida en su conjunto.

[*Los encuestados deben reflejar su respuesta en el cuadernillo del encuestador. Entregar el cuadernillo para que respondan a esta pregunta, recogerlo y continuar con C17*].

Máxima satisfacción
imaginable con su vida



Mínima satisfacción
imaginable con su vida

Su nivel de satisfacción global con su vida es:

C17. En una escala de 1 (muy fácil) a 10 (muy difícil), ¿podría indicar el grado de dificultad que ha encontrado a la hora de responder a este cuestionario?

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
1 2 3 4 5 6 7 8 9 10

Comentarios/Sugerencias:

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

MUCHAS GRACIAS POR SU COLABORACIÓN

TARJETAS

TARJETA 1

- **Cuidado personal**: asearlo, acompañarlo a ir al aseo; cambiarle el pañal; vestirlo/desvestirlo; darle de comer y de beber; ayudarlo a sentarse, a acostarse y a ponerse en pie.
- **Movilidad**: ayudarlo a desplazarse y a caminar dentro y fuera de casa.
- **Tareas domésticas**: cocinar, limpiar, fregar, planchar, pequeños arreglos.
- **Actividades prácticas**: tomar la medicación; gestionar el dinero; utilizar el teléfono; acompañarlo a ir a comprar, a coger el transporte público y en sus visitas al médico, a familiares y a amigos.

TARJETA 2

- Necesita que le den de comer y de beber.
- Tiene incontinencia urinaria y fecal y necesita que otra persona lo limpie y lo cambie.
- Es incapaz de realizar la mayoría de las actividades de cuidado personal (bañarse, vestirse, lavarse, peinarse, afeitarse, etc.). Necesita que alguien lo reemplace en la actividad.
- Necesita ayuda para desplazarse fuera y dentro del hogar y/o ponerse de pie.
- Es incapaz de realizar la mayor parte de las tareas del hogar (limpieza, comida, etc.), debido a sus problemas de salud.
- Tiene dificultades para vivir solo/a. Necesita ayuda para manejar el dinero y las medicinas, así como para tomar decisiones en algunas de las actividades de la vida diaria (pero no en todas). No opone agresividad (ni física ni verbal) a la ayuda que recibe.

TARJETA 3

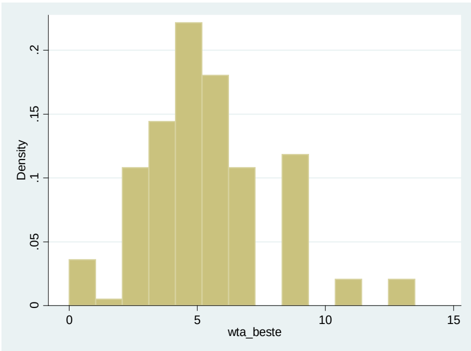
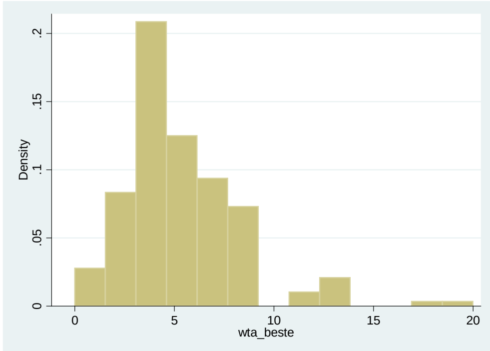
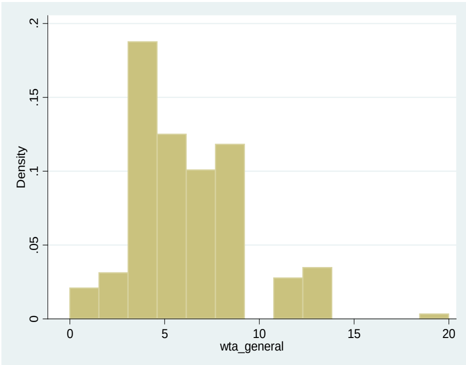
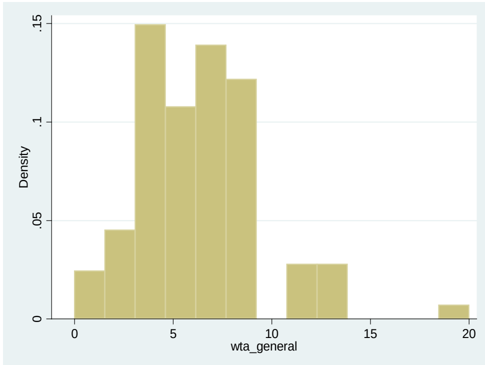
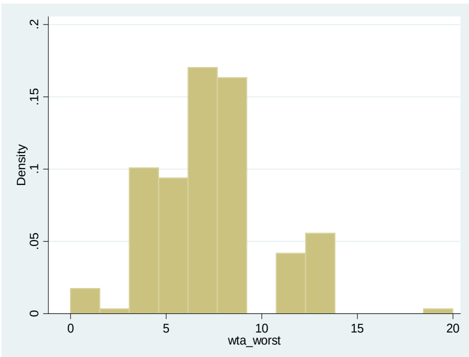
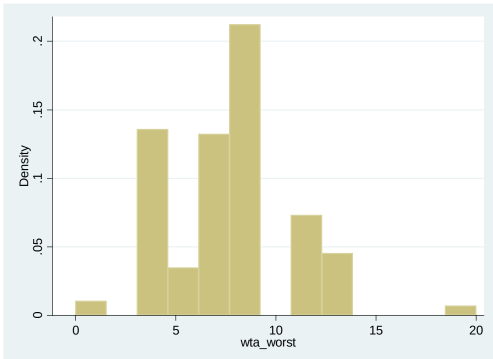
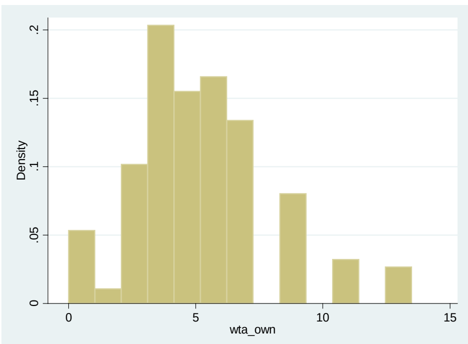
Número de horas diarias:

- **3 horas** ayudándolo en su **cuidado personal**: asearlo, acompañarlo a ir al aseo; cambiarle el pañal; vestirlo/desvestirlo; darle de comer y de beber; ayudarlo a sentarse, a acostarse y a ponerse en pie.
- **1 hora** para ayudarlo a **desplazarse** y a caminar dentro y fuera de casa.
- **3 horas** ayudándolo a realizar las **tareas domésticas**: cocinar, limpiar, fregar, planchar, pequeños arreglos.
- **1 hora** ayudándolo a realizar **actividades prácticas**: tomar la medicación; gestionar el dinero; utilizar el teléfono; acompañarle a ir a comprar, a coger el transporte público y en sus visitas al médico, a familiares y a amigos.

TOTAL: 8 HORAS DIARIAS

3.5. APPENDIX 3B

Figure 3B.1. WTA values (€/hour/day)

	Carers	Non-carers
WTA_{best}		
$WTA_{general}$		
WTA_{worst}		
WTA_{own}		

CHAPTER 4

TESTING THE EFFECT OF LEARNING ON PREFERENCE REVERSALS

ABSTRACT

The main objective of the study presented in this chapter is to test whether the frequency of preference reversal (PR) between choice and valuation decisions declines as subjects learn from the acquisition of experience with those tasks. An experiment was conducted over two independent sessions, in order to distinguish the effect of the mere repetition of tasks (first session) from the effect of repetition in combination with feedback on the consequences of decisions made (second session). Participants were 319 undergraduates at the University of Murcia (Spain), who were split into three groups, according to the nature of the lottery outcomes: 'money' (Group 1), 'years of life in good health' (Group 2) and 'days without back pain' (Group 3). The typical asymmetric pattern of PR was found, with the proportion of standard preference reversal (SPR) being overwhelmingly higher than the rate of non-standard preference reversal (NSPR): 50% *vs.* 1.5% and 53% *vs.* 2.9% in sessions 1 and 2, respectively. Furthermore, respondents in the two groups where the lotteries offered health outcomes (especially those who were exposed to years of life) were more likely to incur in SPR than those who dealt with monetary outcomes, as well as women as compared with men. The repetition of tasks in combination with feedback contributed significantly to attenuate the frequency of SPR, unlike the repetition of tasks on its own. This led to a significant improvement in the number of consistent responses over the second session, although this improvement was limited due to the rise in the frequency of NSPR over that session.

4.1. INTRODUCTION

Procedural invariance –the assumption that preferences are independent of the method used to elicit them– is taken for granted in standard economic theory (Tversky *et al.*, 1990). For instance, if a subject prefers option A over option B in a straight choice, in order to be consistent, he/she should value A more highly than B in separate valuation tasks, as well as to prefer A over B in a ranking exercise (Oliver, 2013a). However, a vast body of research has found that normatively equivalent preference elicitation methods frequently lead to different preference orderings, thus violating procedural invariance and raising the fundamental question of which method best reflects the individual's underlying preferences (Braga and Starmer, 2005; Oliver, 2006). Failures of procedural invariance have serious implications for decision making grounded on preference-based methods (such as cost-benefit analysis and cost-utility analysis) and for the foundations of economic theory (Cubitt *et al.*, 2004). Experimental studies have identified different violations of procedural invariance, including the preference reversal (henceforth PR) phenomenon and the disparity between willingness to accept (WTA) and willingness to pay (WTP)[1], among others.

In this chapter we will confine our attention to PR, with a special emphasis on the role that the acquisition of experience has on this behavioural anomaly. To set the background for the study, the remaining of the Introduction first classifies the different forms of PR identified in the literature, including an overview of studies that have focused on PR in relation to health outcomes; then it outlines the mains causes of PR; and, finally, it summarises the evidence on the effect of learning on PR.

4.1.1. Taxonomy of preference reversals

The most well-known example of PR was first reported by psychologists (Lichtenstein and Slovic, 1971; Lindman, 1971; Lichtenstein and Slovic, 1973; Slovic, 1975). This case involves two binary lotteries[2] with similar expected values. One of them (called the 'P-bet' or

---

[1] WTP and WTA are two contingent valuation techniques that are used to elicit maximum buying prices and minimum selling prices, respectively. Under conventional economic theory, WTP and WTA measures for the same commodity should be approximately the same (Willig, 1976). However, a large number of studies have found WTA values to surpass by far WTP values (Knetsch and Sinden, 1984; Kahneman *et al.*, 1990). The disparity between WTP and WTA is particularly notable when the good under valuation represents a significant part of consumer's budget and the associated transactions costs are large (Brown and Gregory, 1999).

[2] Throughout this chapter, the terms 'lottery', 'gamble' and 'bet' are used interchangeably.

'P', for short) offers a relatively high probability of a large prize, whilst the other (the '$-bet or 'S', for short) offers a smaller chance of a larger prize.[3] Subjects are asked to value both bets separately by placing monetary values on them –usually elicited as minimum selling prices– and to make a straight choice between the two gambles. Most subjects choose P but value $ more highly.[4] This particular inconsistency is known as standard preference reversal (henceforth SPR). The opposite pattern (that is, when $ is chosen but P is valued more highly) is called non-standard preference reversal (hereafter NSPR) and is seldom observed. The asymmetry between the two types of PR (SPR and NSPR) suggests that the presence of PR cannot be explained by random error alone (Braga and Starmer, 2005).[5]

Aside from this classic example, other forms of PR have been identified in the literature, which could be classified into three broad categories, depending on the preference elicitation procedures (e.g. pricing, matching, choice, rating, ranking)[6] involved:

a) Choice-judgment reversals: This category includes the classic example of PR that has been previously explained (also known as choice-pricing reversal), which has been found with both selling and buying prices (Loomes *et al.*, 2010). Similar types of reversals are the discrepancy between choice and matching[7] (Tversky *et al.*, 1988; Slovic *et al.*, 1990; Fischer *et al.*, 1999) and between choice and rating (Goldstein and Einhorn, 1987; Tversky *et al.*, 1990; Fischer and Hawkins, 1993).

b) Judgment-judgment reversals: Several forms of judgment-judgment reversals have been found, including the rating-pricing discrepancy (Schkade and Johnson, 1989; Fischer and Hawkins, 1993) and the matching-matching discrepancy –either using two simple matching tasks (Slovic *et al.*, 1990) or a double (or chained) matching (Delquié, 1993). Two particular cases of matching-matching reversals are the divergence between the certainty equivalence (CE) and the probability equivalence (PE) methods and between the value lottery equivalence (VLE)

---

[3] Sometimes the P-bet and the $-bet are also labelled 'H-bet' and 'L-bet', denoting high probability and low probability, respectively.

[4] For instance, in the experiments conducted by Lichtenstein and Slovic (1971), between 51-83% of respondents chose P but placed a higher selling price on $.

[5] In this regard, there is a nascent literature around modelling decision errors and/or preferences imprecision (Hey, 2005; Loomes, 2005).

[6] Pricing and matching are two types of valuation techniques, in that in both of them subjects are enquired to attach a number (a value) to an object, depending on their preferences. Moreover, pricing, matching and rating reflect 'judgments'.

[7] In a matching task, two alternatives are presented, one of which has some missing information (e.g. the probability or the payoff). Respondents are asked to specify the value of the missing parameter that would make the two options be equally attractive for them.

and the probability lottery equivalence (PLE) methods. A number of studies have found higher utility values with the PE than with the CE (Hershey and Schoemaker, 1985), on the one hand, and with the PLE as compared with the VLE (Delquié, 1993), on the other hand.

c) Choice-choice or choice-ranking reversals: Although scarcer, there are also some examples of reversals between two choices (Tversky and Kahneman, 1986) and between a choice and a similar task, such as a ranking exercise (Bateman *et al.*, 2007).

The aforementioned examples refer to non-health related outcomes, mainly money. An overview of the different forms of PR that have been observed using health outcomes is presented below.

a) Choice-judgment reversals

Oliver (2013a) provides the closest replication of the classic example of PR regarding health outcomes that has been reported until now. His study comprised two separate experiments, which had basically the same design. In both cases, respondents were asked to imagine that they had a treatable health condition (which was described as an EQ-5D health state) and were presented with two treatments. Both interventions were defined as lotteries, one of them being a $-bet (modest probability of being in perfect health) and the other a P-bet (higher probability of a modest outcome).The outcomes of both gambles were also characterised as EQ-5D states. Next, respondents were required to make three types of tasks: a direct choice between the two treatments and two valuation tasks (in the form of WTP). One of the WTP questions (the 'open' valuation) had an open-ended format, whereas the other (the 'assisted' valuation) offered respondents a payment card. The author conjectured that PR would be less frequent when comparing the choice with the assisted valuation than when comparing the choice with the open valuation. This hypothesis was rejected because the frequency of PR did not differ significantly between the two formats. Specifically, with the open valuation procedure, the rate of PR (considering strict and weak PR as a whole)[8] ranged from 35% (in the second experiment) to 37.5% (in the first one). With the assisted valuation procedure, 37.5% and 36.8% of subjects exhibited PR in experiments 1 and 2, respectively. Thus, considering both weak and strong PR, the rates of inconsistent responses found in this study were quite

[8] In this study, indifference was allowed in both the choice and valuation tasks. A 'strict PR' occurred when P was chosen and $ was valued more highly (SPR) or vice versa (NSPR). A 'weak PR' happened when respondents expressed indifference between P and $ either in the choice or in the valuation task (but not in both cases).

substantial. The main difference between the results of the two experiments is that strict and weak PR were almost equally frequent in the first experiment (with both valuation procedures), whereas in the second experiment weak PR was much more frequent than strict PR. Perhaps the most striking result of this study is that SPR was rarely observed (it only accounted for 2-5% of responses), being even less common than NSPR in experiment 1 (15-18% of answers). It is also worth noting that the analysis of the explanations of respondents' answers revealed that many of them used heuristics, which cast doubts about the reliability of their stated preferences.

Another recent example of reversals between choice and valuation procedures can be found in Oliver (2013b). This study also consisted of two independent experiments. In both cases it was tested whether PR is influenced by the frame (or perspective) under which the preference elicitation tasks were made, distinguishing between a personal frame (where respondents made decisions that affected themselves as patients) and a social frame (which entailed making decisions for others). Additionally, in the second experiment the effect of incentives on PR was also examined. In the two experiments strict PR (that is, SPR + NSPR) was more prevalent under the social context (43.3% and 31.7% of responses in experiments 1 and 2, respectively) than under the personal one (36.7% and 18.4%, respectively). The divergence between the two perspectives is explained by the higher rate of SPR in the social context than in the personal one, since NSPR was almost equally frequent under both frames. In addition, the difference between the two frames in terms of the proportion of strict PR was more pronounced with incentives (experiment 2) than without incentives. According to Oliver, the higher rate of strict PR in the social frame than in the personal one is mainly ascribable to the fact that, in the choice task, respondents displayed a stronger risk aversion (and, in consequence, a greater preference for the P-bet over the $-bet when choosing) under the social context than under the personal one. This means that respondents were more risk averse when they had to choose for others than when they had to choose for themselves. Conversely, the valuations were rather insensitive to both the decision making frame and the presence of incentives.

On the other hand, there is also some evidence of choice-matching reversals with health outcomes. For example, Sumner and Nease (2001) conducted an experiment in which subjects were asked to choose between two hypothetical clinical scenarios and to match both alternatives. The proportion of subjects who chose the alternative A (the milder health state) was higher than the proportion of those who favoured A in the matching task. Conversely, the proportion of respondents who favoured the alternative B (the most severe state) in the matching task was higher than the percentage of those who chose B. The authors surmised that their findings could be attributable to the prominence effect. Another specific case of choice-matching discrepancy relative to health is the so-called 'maximum endurable time' (MET) phenomenon (Sutherland *et al.*, 1982). The MET is considered to be that period of time beyond which shorter lifespan is preferred to a longer one. For instance, Stalmeier and Wakker (1997) reported that many respondents preferred to live 10 years suffering 5 days of migraine per week rather than to live 20 years in that health state. However, when the same individuals were asked to equate both alternatives with a shorter period of time in good health –using the time trade-off (TTO) technique–, they generally demanded more healthy life years for the longer duration with migraine than for the shorter one. Similar results have been found in other studies, both with the TTO (Unic *et al.*, 1998; Stalmeier *et al.*, 2001) and with the standard gamble (SG) methods (Stalmeier and Verheijen, 2012).

Other type of choice-judgment reversal observed in the health economics literature is the discrepancy between WTP values and ordinal ranking (considering that a ranking resembles a choice task). In those studies in which WTP is elicited for various health care programmes, a basic consistency test consists of comparing the explicit ordinal ranking of the different programmes with the implicit ranking derived from their WTP values. For instance, this test has been reported in the studies undertaken in the framework of the 'EuroWill' project[9] (Olsen and Donaldson, 1998; Gyldmark and Morrison, 2001; Shackley and Donaldson, 2002). Overall, only 21-43% of respondents in these studies provided WTP values for health care programmes that were consistent with the explicit ordinal rankings of the programmes.

b) Judgment-judgment reversals

Within-method inconsistencies (also known as 'internal' inconsistencies) are a form of judgment-judgment reversal that has been extensively examined in the context of health state utility assessment. This kind of reversal takes place when logically equivalent variants of the same procedure lead to different utility scores. Some examples can be found in Llewellyn-Thomas *et al.* (1982), Rutten-van Mölken *et al.* (1995), Bleichrodt (2001), Oliver (2003), and Pinto and Abellán (2005). All these studies compare the utility score for a given health state elicited by using one-stage measurements (e.g. a simple SG question) with the one elicited by using two-stage or 'chained' measurements (e.g. chaining two SG questions linked each other by means of a common outcome). The general conclusion that stems from these papers is that the fact of replacing the original failure outcome by another failure outcome generally produces

[9] The EuroWill (Donaldson, 1999) is a research project funded by the European Commission. Its main objective was to investigate the feasibility of using WTP as a method to set priorities in different health care systems. As part of this project, WTP surveys were conducted in six European countries: Denmark, France, Norway, Portugal, the Republic of Ireland and the United Kingdom.

higher utility values than the conventional SG utility scores. Therefore, these studies raise questions about the internal consistency of the SG (at least under expected utility).

On the other hand, Bleichrodt *et al.* (2001) focused on the discrepancy between the CE and the PE methods using years of life as outcomes. They found the typical disparity (i.e. PE yielded higher utilities than CE). In a subsequent research, Bleichrodt *et al.* (2007) extended the previous analysis to three additional utility elicitation methods, namely, the value equivalence (VE), probability lottery equivalence (PLE) and value lottery equivalence (VLE) methods.[10] PE, CE and VE are riskless-risk methods, meaning that only one of the two alternatives presented involves risk. On the other hand, VLE and PLE techniques are risk-risk methods, since the two options presented are risky. Under expected utility theory, it was rejected that these methods lead to the same results. The comparison among the utilities generated by the five methods revealed the following general ordering: VE > PE > CE > PLE > VLE. Therefore, riskless-risk methods gave rise to higher utilities than risk-risk methods. Furthermore, the observed discrepancies were more pronounced between the riskless-risk methods than when comparing the two risk-risk methods. Indeed, the utilities elicited by PLE and VLE did not differ significantly.

There are also examples of reversals between matching and rating in some studies that have compared death with severe health states. For instance, Robinson *et al.* (1997) observed that most subjects rated the presented health states higher than death on the visual analogue scale (VAS). However, when the same subjects were asked to choose (in a TTO question) between living 10 years in these health states and immediate death, many preferred the latter option. Overall, 67.4% of respondents scored at least one state above death on the VAS, but then rated the same state as worse than death in the TTO, thus incurring in PR. The authors of this study suggested that, to a large extent, this preference inconsistency might be explained by a neglect of the duration of the health states in the VAS task, along with the existence of a MET, putting death at the bottom of the VAS. Conversely, Robinson *et al.* (2001) obtained a similar preference ordering with the VAS and with the SG.

[10] The CE method elicits the duration that makes a subject be indifferent between an outcome for certain and a risky lottery, while the PE method elicits the probability at which an individual is indifferent between an outcome for certain and a risky lottery. The VE method elicits the duration that makes a subject indifferent between two alternatives, one of which is riskless and the other is risky. On the other hand, the difference between the VLE and the PLE methods is that the former elicits the duration that makes an individual be indifferent between two risky lotteries, whereas the latter elicits the probability at which the indifference between two risky lotteries is attained.

c) Choice-choice reversals

An example of choice-ranking discrepancy involving health outcomes is the study conducted by Oliver (2006). It should be noted that this study used a 'social' framing rather than a 'personal' one (i.e. the lotteries did not entail individual treatment options). Participants in this experiment were enquired to choose between two countries with different life expectancy distributions, as well as to value both alternatives by attaching a CE value to each one and to make a ranking exercise in which the two hypothetical countries were compared with a number of other hypothetical countries. By defining the different alternatives in terms of life expectancy distributions, respondents were implicitly encouraged to consider the implications for themselves and for other people. The comparison between choices and CE values revealed a strong asymmetric pattern of PR: 37% of subjects incurred in SPR and nobody exhibited the opposite pattern (NSPR). When comparing choices and rankings, the frequency of SPR dropped by half and the asymmetry between SPR and NSPR was less pronounced, but it still remained. To be exact, 14% and 6% of subjects committed SPR and NSPR, respectively. It has been speculated that the lower frequency of PR with ranking as compared with conventional valuation tasks could be due to the fact that, in the former case, subjects are presented with more ample information and, therefore, their responses are expected to reflect a more thorough reflection of the alternatives presented. Consequently, the $-bets are less likely to be overvalued in a ranking exercise than in a valuation task.

Bleichrodt and Pinto (2009) also found a choice-ranking discrepancy, specifically, when comparing a choice between two risky treatments with a ranking of health states (including death). Strikingly, the reversal persisted even when the ranking of health states was derived from choices. Therefore, this latter form or reversal cannot be explained by the use of different information strategies, because it resulted from the comparison between two choice-based tasks (a pairwise choice between the two treatments and a choice-based ranking).

4.1.2. Explanations for preference reversals

In an attempt to explain the PR phenomenon, several causes have been put forward (for a review, see Seidl (2002). The most widely accepted explanations are those that see PR as a consequence of the overpricing[11] of the $-bet. According to Tversky *et al.* (1990), this is the

[11] A lottery is said to be overpriced when a subject prefers its elicited price over that bet in a direct choice between the price and the bet. By contrast, a lottery is said to be underpriced when an individual prefers the bet over its price in a direct choice (Tversky *et al.*, 1990).

major source of PR. To be exact, this overpricing explains about 65% of PR, while the underpricing of the P-bet only accounts for around 6% of PR (as a whole, the overpricing of $ and/or the underpricing of P account for approximately 90% of PR).

Slovic and Lichtenstein (1968) provided an explanation for the overpricing of the $-bet. They observed that choices and ratings of attractiveness were primarily determined by the probabilities of the gambles, whereas prices (bids) were predominantly correlated with the payoffs (the amounts to win or to lose). These researchers conjectured that the differences observed between choice and pricing tasks reflected the use of different information strategies (e.g. heuristics) in each case. On the one hand, when making a choice, each attribute of one bet can be directly compared with the same attribute of the other bet, and there is no natural starting point. On the other hand, when valuing a gamble, there is a clear starting point, which is the amount to win (people tend to focus on the outcome attribute as starting value and to adjust it downwards by taking into account the other attributes of the lottery). Since the P-bet offers a high probability of winning, its elicited value is adjusted downwards only slightly. Conversely, the $-bet requires a substantial downward adjustment to account for the small chance of success. As a result, when subjects value the $-bet, they often anchor on its best outcome, but fail to adjust the overall value of this gamble sufficiently downwards to take account of other attributes (Lichtenstein and Slovic, 1971). This, consequently, causes an overpricing of the $-bet and a mismatch between lottery choices and certainty equivalent values, giving rise to a reversal of preferences (Tversky and Thaler, 1990).

The view that PR is related to the use of different information processes in the choice and valuation tasks has derived in two formal explanations: the prominence effect hypothesis (Tversky *et al.*, 1988; Slovic *et al.*, 1990) and the compatibility hypothesis (Tversky *et al.*, 1988; Slovic *et al.*, 1990; Tversky *et al.*, 1990; Slovic, 1995). According to Slovic *et al.* (1990), the PR phenomenon is a consequence of the combination of both effects. However, other authors have found the prominence effect to be much stronger than the compatibility effect (Fischer and Hawkins, 1993).

The prominence effect hypothesis presumes that there is an attribute which, in some sense, is the most important (or prominent) to the subject, and that this attribute weights more heavily in choice than in matching (Tversky *et al.*, 2000). Given that individuals seems to perceive probability as more important than payoffs –as suggested by the fact that the rating of bets is dominated by probability (Slovic *et al.*, 1990)–, the prominence effect would make the P-bet be relatively more attractive than the $-bet in a choice task, since the former offers a higher probability of success than the latter. A possible explanation for the prominence effect is that choice requires more qualitative reasoning, such as the use of a lexicographic strategy (e.g. to select the alternative that is ordinally superior on the most prominent attribute). Matching, by contrast, requires a more quantitative assessment (Tversky *et al.*, 1988).

On the other hand, according to the scale compatibility hypothesis, the weight of any attribute (e.g. probability, payoff) is heightened when it is compatible with the scale in which the responses are expressed. Specifically, given that the price for a lottery is expressed in the same unit as the payoffs (i.e. in monetary terms), the payoffs are likely to be weighted more heavily in pricing than in choice tasks. In addition, since the payoffs are much larger in the $-bet than in the P-bet, the main consequence of a compatibility bias is the overpricing of the $-bet, which, as mentioned earlier, is considered to be the major cause of PR. Consequently, the use of non-monetary outcomes should lower the frequency of PR (Tversky *et al.*, 1990), insofar no compatibility between outcomes and prices takes place in this case. The findings reported by Slovic *et al*. (1990) support this conjecture. These researchers undertook an experiment involving six pairs of bets, half of which involved monetary payoffs and the other half entailed different types of non-monetary gambles, such as a one-week pass for the theatre or a dinner for two people. They found that the proportion of SPR was significantly lower with the non-monetary bets (24%) than with the monetary bets (41%). The authors attributed this finding not only to the compatibility effect, but also to the prominence effect.

4.1.3. The effect of learning on preference reversals

It is important to note that much of the empirical research on PR (and other preference anomalies) comes from one-shot experiments (i.e. experiments where there is no repetition of tasks) conducted in non-market settings. Some researchers have purported that the inconsistencies detected in these settings could be attributable to unfamiliarity (or inexperience) with the tasks performed. For instance, Binmore (1994, 1999) argued that preference anomalies should be a matter of concern for economists only if they persist even after individuals have repeated the same tasks several times, have received feedback on the consequences of their decisions and have been incentivised to think carefully about them. Along these lines, there is some evidence that preference anomalies tend to diminish in certain experimental markets where repetition, feedback and real incentives conflate (Coursey *et al.*, 1987; Cox and Grether, 1996; Braga *et al.*, 2009). This fact is consistent with the ‘discovered preference hypothesis’ (Plott, 1996) and the ‘refining hypothesis’ (Loomes *et al.*, 2003). On the whole, these hypotheses assume that individuals have stable and rational (i.e. anomaly-free) preferences.

Nonetheless, when people cope with unfamiliar tasks, their stated preferences are likely to be biased, in the sense that they possibly do not reflect their true underlying preferences. But the participation in repeated markets may contribute to refine individuals' stated preferences, approaching them to their true preferences. Although neither the discovered preference hypothesis nor the refining hypothesis make explicit the factors that promote the discovery of true preferences, it can be inferred from both hypotheses that the aforementioned conditions (i.e. repetition, feedback and real incentives) can each contribute to that discovery (Loomes *et al.*, 2003). Through repetition subjects become more familiar with the tasks they are asked to make and can learn several aspects (e.g. rules) about them. Through feedback, they experience the consequences of their own decisions. If, in addition, real incentives (e.g. money) are provided, those consequences turn out to be for real. Thus, in this view, incentives play a key role because, without incentives, decisions are merely hypothetical (i.e. have no real consequences) and, as a result, learning may be absent (Braga and Starmer, 2005). If this is true, it would entail a limitation for the elicitation of preferences over non-market goods (e.g. health, the environment, education, etc.) because no real incentives can be provided in these cases. Notwithstanding, it must be clarified that the effect of real incentives on PR remains ambiguous (Seidl, 2002). For example, similar results have been found in one-shot experiments both with and without incentives (Lichtenstein and Slovic, 1971; Grether and Plott, 1979). In general, compared with the use of hypothetical incentives, real incentives tend to increase risk aversion (Smith and Walker, 1993), although this effect seems to be small when the tasks are cognitively easy (e.g. choices) (Camerer and Hogarth, 1999).

In sum, the discovery of preferences can be regarded as a process of learning to avoid errors and about one's actual preferences (Braga and Starmer, 2005). It should be noted that each repetition, feedback and real incentives can be found in both market and non-market environments. But, interestingly, the evidence shows that non-market institutions that involve these three factors are less effective in eroding anomalies than market institutions in which the same three conditions are met. Cox and Grether (1996) offered an example of this kind. They conducted an experiment involving a series of repeated valuations (five repetitions for each the P-bet and the $-bet) and a single choice (to be made at the end of the experiment). After each valuation, respondents received feedback on the consequences of their decisions. Different incentive-compatible mechanisms were used, such as the second-price Vickrey auction[12]

[12] The SPA is a specific case of Vickrey auction in which the seller who is willing to sell at the lowest price sells the auction at a price equal to the second lowest selling price. Similarly, when the SPA is used to elicit buying prices, the buyer who is willing to pay the highest price for the good buys it at the price of the second highest bid (Knetsch *et al.*, 2001).

(henceforth SPA) and the Becker, DeGroot, Marschak (BDM) variant of the Vickrey mechanism[13] (Becker *et al.*, 1964). Participants had to deal with only one of these mechanisms. The comparison between the first valuations and the single choice revealed the common asymmetric pattern of PR (i.e. SPR was far more frequent than NSPR) with both the SPA and the BDM schemes. Comparing the final valuations with the choice, the same pattern still remained with the BDM, while SPR and NSPR were almost equally frequent with the SPA (giving rise to a symmetric pattern). However, by the end of the experiment, the overall frequency of PR was higher with the SPA than with the BDM. The different patterns observed with these two mechanisms are somewhat odd because the BDM resembles a SPA in which a subject competes against another agent who bids at random (Braga *et al.*, 2009). Since the SPA is a market institution whilst the BDM is not, it might be hypothesised that the differences between the two procedures are explained because of the specific features that are unique to a market environment (e.g. the presence of actual competitors, the establishment of the market price via equilibrium between supply and demand, the rules being used, etc.).

Nonetheless, it must be stressed that even theoretically equivalent market institutions can also differ in their propensity to reduce inconsistencies. This point is illustrated in the study undertaken by Braga *et al.* (2009). In the first experiment of this study, both the SPA and another variant of the Vickrey auction were used, to be exact, the second-to-last price auction (henceforth StLPA)[14]. At the beginning of this experiment, both procedures led to the typical strong asymmetry between SPR and NSPR. But, after a few rounds, clear differences emerged between the two mechanisms. In particular, by the fifth round of valuations, SPR was significantly less frequent with the SPA, whereas the frequency of NSPR had quadrupled the initial values. As a result of these two opposite trends, both types of PR were almost equally frequent by the fifth round, thus leading to a symmetric pattern of PR. With the StLPA the frequency of SPR also declined, while NSPR experienced an upward trend. However, since the change between the first and fifth rounds was not significant, SPR was still significantly more frequent than NSPR by the fifth round. The different pattern found between the two institutions used in this experiment is related to the fact that the valuations for the $-bets rose relatively quickly with the SPA, whilst they fell relatively slowly with the StLPA.

[13] In the BDM mechanism each subject is endowed with a good and states a minimum selling price for it. Next, a random offer is drawn for the good. If that offer is equal to or greater than the stated minimum selling price, the individual sells the good and receives the random offer. By contrast, if the random offer is lower than the stated minimum selling price, there is no trade and, therefore, the subject retains the good.

[14] Unlike in the SPA (where the lowest bidder sells at the second lowest price), in the StLPA all but the two highest bidders sell at a price equal to the second highest bid (Braga *et al.*, 2009).

Furthermore, some anomalies can decrease with a specific mechanism but, in turn, new biases can arise with the same procedure. This is exactly what Braga *et al.* (2009) found in their second experiment. In this case, although SPR diminished in presence of ongoing feedback, beyond a certain number of valuation rounds (to be exact after four repetitions), NSPR became more frequent than SPR. Consequently, while the typical asymmetric pattern of PR had disappeared by round 5, a non-standard asymmetry appeared beyond round 7 (i.e. NSPR became more frequent than SPR). For that reason, the authors concluded that market experience erodes some anomalies (i.e. SPR), but it creates new ones (i.e. NSPR).

It is worth noting that, in the experiment just mentioned, respondents who received feedback behaved differently depending on the type of feedback and the information that it conveyed. In particular, the subsidence of PR in this experiment was explained to a large extent by the direct experience of losses. To be exact, subjects who received round-by-round feedback on lottery outcomes tended to reduce their selling prices for the \$-bets after a round in which they had failed to sell a \$-bet and lost (i.e. after experiencing a direct loss). On the contrary, those who sold a \$-bet and learnt that they would have lost if they had kept and played it (i.e. those who experienced an indirect loss effect) behaved similarly to those who received no feedback. A possible explanation for these findings is that subjects who failed to sell the \$-bets and lost might feel induced to adjust their \$-valuations downwards in subsequent rounds with the aim of avoiding new losses. Conversely, those who sold and won might have no incentives to revise their valuations (Braga and Starmer, 2005). Given that the indirect loss experience effect was not as strong as the direct loss experience effect, Braga *et al.* (2009) concluded that the direct effect could not be merely a consequence of learning about probabilities. This interpretation is consistent with the 'market discipline hypothesis' (Loomes *et al.*, 2003), which puts forward that subjects tend to correct their prior errors if they have proved to be costly. When eliciting minimum selling prices –such as in the experiment by Braga *et al*. (2009)–, respondents make a costly mistake only if the market price is comprised between the stated value and the true underlying value (i.e. if stated value > price > true value). In that situation, a subject would fail to sell at an attractive price, thus committing a costly mistake (Braga and Starmer, 2005). It is important to note that the market discipline hypothesis was formulated in the context of a market in which individuals are not money-pumped when their preferences are inconsistent. Therefore, here, a 'costly mistake' does not mean that subjects are sanctioned or punished. Unsurprisingly, there is evidence that inconsistencies quickly vanish when individuals are exposed to money pumps (Chu and Chu, 1990).

In short, four main conclusions can be drawn from the above-mentioned studies. First, the mere participation in repeated experimental markets does not suffice to eliminate preference anomalies, inasmuch they are robust to some institutions whereas they subside in others. Second, the dependence-mechanism illustrated in these studies suggests that learning is driven by different factors that are contingent upon the elicitation mechanism used. Third, learning can erode some anomalies, whilst others can be exacerbated. Fourth, individuals' reactions to feedback seem to depend on whether it informs about direct or indirect consequences and on whether those consequences are positive or negative.

4.1.4. Study objectives

The experiment reported below adds to the literature on the effect of learning on the consistency of stated preferences. The primary aim of the present study is to examine whether PR (in relation to both health and monetary outcomes) between choice and valuation tasks can be attenuated by fostering learning through the acquisition of experience with those tasks. In this respect, two learning effects are differentiated: the effect of the mere repetition of tasks and the combined effect of repetition and feedback on the consequences of decisions made. Other additional objectives of this study are to investigate whether the type of outcomes (health, money), the starting point presented in the valuation tasks and gender have an influence in the rate of PR. The main contributions of this study are that, to the best of our knowledge, it is the first one to test the effect of learning on PR in relation to health outcomes, as well as the first one to compare differences in the proportion of PR between health and monetary outcomes.

4.2. METHODS

4.2.1. Study design

4.2.1.1. Overview

The study was conducted in two experimental sessions separated by one week. Prior to the actual experiment, a pilot study (with a sample of 10 master-level economics students) was undertaken to test the questionnaires. Participants were a convenience sample of 319 undergraduates at the University of Murcia (Spain), being most of them students at the Faculty of Economics and Business. The sample was recruited by posting advertisements in different locations throughout the campus of that university. The same subjects who participated in the first session also engaged in the second one. The sample was split into three main groups and subjects were randomly allocated to one of them. The three groups differed in the kind of outcomes they were presented with: ‘money’ (Group 1), ‘years of life in good health’ (Group 2) and ‘days without back pain’ (Group 3), considering the ‘success’ outcomes of the gambles. The corresponding failure outcomes in Groups 1, 2 and 3 were ‘€0’, ‘death’ and ‘days with back pain’, respectively. All groups were further divided into three subgroups, as it will be explained below.

Table 4.1 summarises the distribution of the sample by groups, along with the main demographic characteristics (age and gender distribution) of the sample and of each group. Groups 2 and 3 were almost equally-sized (n = 78 and 81, respectively), whilst Group 1 (n = 159) was composed of the same number of subjects than the other two groups together. The three groups were quite homogeneous in terms of age, with a mean age around 22 years in the three cases. Women (58.9% of the whole sample) outnumbered men in all groups, especially in Group 3.

Table 4.1. Demographic characteristics and distribution of the sample

	All	Group 1 (‘Money’)	Group 2 (‘Years of life’)	Group 3 (‘Days free of back pain’)
N (%)	319	159 (49.8)	79 (24.8)	81 (25.4)
Mean (SD) age	22.2 (4.9)	21.8 (3.8)	22.7 (4.9)	22.5 (6.7)
Women (%)	58.9	56.6	57.0	65.4

SD: Standard deviation.

Different questionnaires were made for each group (see Appendix 4A for details). The two sessions took place in a computer room at the Faculty of Economics and Business of the University of Murcia, under the supervision of members of the research team. The questionnaires opened stating that the motivation of the study was to get a deeper insight into individual preferences under uncertainty. The introduction also outlined the main tasks that the experiment would involve and encouraged participants to express their true preferences, since there were neither wrong nor right answers. The instructions for each question were displayed on the computer screens. Respondents could answer the questions at their own pace. If they had doubts or problems at any point of the session, they could ask the researchers who were supervising the course of the experiment.

To make the experiment more manageable, several turns were arranged for each session and subjects chose the most convenient turns (one for each session) for them. There was no relationship between the distribution of participants in the three groups and the subjects who engaged in each turn. On average, around 30 subjects participated in each turn. The mean duration was, approximately, 25 minutes in the first session and 20 minutes in the second one.

4.2.1.2. Gambles and tasks

Both experimental sessions involved two types of tasks: straight choices between two gambles (P-bet and $-bet) and separate valuation of each gamble. Two different pairs of lotteries were used, which are outlined in Table 4.2. All bets (X, p) offered X (the best outcome) with probability 'p' and zero (the worst outcome) otherwise. The two paired lotteries were matched or almost matched in their expected values, but their probabilities and outcomes differed to a great extent. One of them (the P-bet) offered a relatively large probability of winning a modest outcome, whereas the other (the $-bet) gave a lower chance of obtaining a larger outcome. Thus, gambles A and C were the P-bets and gambles B and D were the $-bets. The pairs (AB) and (CD) were used in both sessions (the same number of times). It should be noticed that the labels we use in this chapter for the different lotteries were not used in the experiment, where the bets were labelled according to the outcomes used in each group. In Group 1 the gambles were labelled 'Alternative 1' and 'Alternative 2'— in the valuation tasks— and 'Alternative A' and 'Alternative B' —in the choices. In Group 2 and Group 3 the labels 'Treatment 1' and 'Treatment 2' and 'Drug 1' and 'Drug 2', respectively, were used.

Table 4.2. Lotteries used in the experiment

	Pair 1	EV	Pair 2	EV
P-bet (A, C)	A: (8, 0.95)	7.6	C: (5, 0.8)	4.0
$-bet (B, D)	B: (27, 0.3)	8.1	D: (21, 0.2)	4.2

EV: Expected value.

The valuation tasks consisted of eliciting the certainty equivalent (CE) value of the gambles. Specifically, in each CE question, participants were asked to state the amount 'M' for which they would be indifferent between one of the bets presented in Table 4.2 and M for sure (i.e. with a 100% chance). Figure 4B.1 (Appendix 4B) displays a screenshot where it can be seen the way in the CE tasks were presented to respondents on the computer screens.

As mentioned earlier, each of the three groups was split into three subgroups. This subdivision was made according to the starting point presented in the valuation tasks: low starting point, high starting point and open-ended starting point. Hereafter, we will use the labels ‘low’, ‘high’ and ‘open’, respectively, in reference to these three formats. Subjects who dealt with the low starting point first observed the worst outcome (zero in all bets) and then they had to vary it until being indifferent between the bet and M. By contrast, subjects who were assigned to the high starting point were given the best outcome (e.g. 8 for lottery A) as starting value and had to vary it until they regarded the bet equally attractive as M. Finally, the open-ended starting point entailed placing the CE value on each gamble directly, since no starting value was provided. In all cases, respondents could increase or reduce their valuations by using the up and down arrow keys of the keyboard. Participants were randomly allocated across starting point formats, seeking a balanced distribution. Overall, 108, 105 and 106 subjects were assigned to the low, high and open starting point formats, respectively. Within a particular group, the number of subjects who dealt with each format ranged from 24 to 28, depending on the group.

With regard to the choices, respondents simply had to select their most preferred option by clicking on it. Indifferent responses were not allowed in the choices. This restriction is commonly imposed in choice experiments, in order to avoid a large number of false indifferent answers (Tversky *et al.*, 1990; Braga *et al.*, 2009). In all choices, the alternative 1 was the P-bet in Group 1 and the $-bet in the other two groups. A screenshot of a choice task is shown in Figure 4B.2 (Appendix 4B).

4.2.1.3. Structure of the sessions

The two experimental sessions had a different structure, as can be seen in Table 4.3. The rationale for doing two sessions was to distinguish the learning effect that is generated by the mere repetition of tasks from the learning effect that is induced by both repetition and feedback. In the first session the repetition of tasks was the only source of learning, whilst over the second session subjects could learn from both the repetition of tasks and feedback on the consequences of their decisions.

The first session comprised six sets: three sets in which the CE value for each of the four lotteries (CE(A), CE(B), CE(C), CE(D)) was elicited, and three sets involving the choice between the P-bet and the $-bet of each pair (A *vs*. B or C *vs*. D). Therefore, each valuation and choice was made three times in the first session. The order according to which the sets (and the tasks within each set) were posed to the respondents was generated at random. For that reason, the order displayed in Table 4.3 is provided for illustrative purposes only. For example, the first session could have been started with a set of choices (e.g. C *vs*. D and then A *vs*. B, or vice versa), followed by a set of valuations (e.g. CE(B), CE(C), CE(A), CE(D), or any other order), and so on.

Only in Group 1, once respondents had finished all the tasks of the first session, they were explained how their reward for taking part in the study would be determined at the end of the next session. After observing an already made example, they had the opportunity to make an example on their own, which was followed by two questions to check whether they had understood the procedure. Since there was no wrong answer to these questions, we assumed that all respondents in Group 1 understood the functioning of the incentive system properly. This mechanism will be explained in detail in the next subsection.

The second session entailed five sets: four sets of valuations (sets 1-4) and one set of choices (set 5). Sets 1-4 had the same structure, but each one referred to a different lottery, which was determined at random. First, respondents were asked to indicate their CE value for the corresponding gamble. After that, they observed the resolution of risk of the lottery they had just valued. The resolution of risk consisted of a visual example (displayed on the computer screen) that simulated the outcomes of the lottery when playing it 10 consecutive times. Specifically, respondents saw human figures coloured in green and red (representing the success and failure outcomes of the gamble in question, respectively), which were distributed across 20 closed doors, according to the probabilities of success and failure of the different lotteries. Figure 4B.3 (Appendix B) shows a screenshot of the resolution of risk of one of the lotteries. Over the resolution of risk process, respondents could watch on the computer a table displaying how many times the success and failure outcomes of the lottery in question had occurred, as well as the CE value that they had stated for that bet. After observing repeatedly how the uncertainty of the gamble was solved, respondents had to value the same bet again, and so forth. The same scheme (CE question followed by the resolution of risk) was made three times (within a same set) for each lottery. By means of the resolution of risk of the gambles, we intended to teach respondents that their decisions could involve either gains or losses.[15] This might help them become more aware of the meaning of the probabilities, as well as to reconsider their

[15] In this chapter, the term ‘loss’ is used as a synonym of gaining nothing.

decisions (e.g. the value they had previously attached to a lottery) when repeating the same task, which could finally lead to more conscious and better informed decisions.

Finally, set 5 of the second session involved two pairwise choices (first between A and B and then between C and D). Thus, in the second session each lottery was valued three times, while the choice between the two paired bets was made only once.

Table 4.3. Structure of the two experimental sessions (example)

Set	First session	Second session
1	CE(A) CE(B) CE(C) CE(D)	CE(A) Resolution of risk (10 times) CE(A) Resolution of risk (10 times) CE(A) Resolution of risk (10 times)
2	A *vs*. B C *vs*. D	CE(B) Resolution of risk (10 times) CE(B) Resolution of risk (10 times) CE(B) Resolution of risk (10 times)
3	CE(A) CE(B) CE(C) CE(D)	CE(C) Resolution of risk (10 times) CE(C) Resolution of risk (10 times) CE(C) Resolution of risk (10 times)
4	A *vs*. B C *vs*. D	CE(D) Resolution of risk (10 times) CE(D) Resolution of risk (10 times) CE(D) Resolution of risk (10 times)
5	CE(A) CE(B) CE(C) CE(D)	A *vs*. B C *vs*. D
6	A *vs*. B C *vs*. D	

Only in Group 1 the first session ended by explaining the incentive procedure, which was performed at the end of the second session.

4.2.1.4. The incentive procedure

At the end of the second session, all respondents were paid in cash for their participation in the experiment. Subjects in Groups 2 and 3 received €15, whilst those belonging to Group 1 were paid a fixed amount (€10) plus a random reward. In order to determine this latter prize, the random lottery incentive procedure[16] was used at the end of the experiment. This scheme pays subjects according to the outcome of one of the tasks, which is selected at random. In our experiment, only the tasks undertaken in the second session were considered. More specifically, the procedure started by randomly selecting a task among those that each respondent had performed throughout the second session. If the selected task was a choice, the gamble that the subject had chosen in that task was played (i.e. the probabilities of the gamble were solved) and, therefore, the respondent could either gain the success outcome or gain nothing. If the chosen task was a valuation, the Becker, DeGroot, Marschak (hereafter BDM) mechanism (Becker *et al.*, 1964) was used to determine the reward. In this case, the computer drew an 'offer' at random. If that offer was at least as high as the stated CE value that the respondent had reported in the selected task, he/she received the offer (since the CE values for the gambles were elicited as minimum selling prices). Conversely, if the random offer was less than the stated CE value, the gamble was played, in the same way as when the selected task was a choice.

4.2.2. Hypotheses

On the basis of the economic theory and the results of prior experiments on PR (e.g. Tversky *et al.*, 1990; Cox and Grether, 1996; Sumner and Nease, 2001; Loomes *et al.*, 2003; Oliver, 2013a; Van de Kuilen and Wakker, 2006; Braga *et al.*, 2009), we established the hypotheses listed below:

Hypothesis 1. Comparison between SPR and NSPR: Bearing in mind the typical asymmetric pattern between the two types of PR, we hypothesised that SPR would be more likely than NSPR in all groups and in the two sessions.

Hypothesis 2. Trend: Following the discovered preference hypothesis, we expected that PR would experience a downward trend over both sessions. The decline in the frequency of PR

[16] The random lottery incentive procedure is considered to be a useful incentive mechanism because it has a number of advantages (Braga *et al.*, 2009). For instance, it avoids potential income effects that could arise if all tasks were counted for real to determine the reward for participation. Besides, given a fixed budget, it allows larger incentives at the level of the individual tasks. Moreover, tests of this scheme have been favourable (Starmer and Sugden, 1991; Cubitt *et al.*, 1998).

over the first session would be explained by the effect of repetition alone, while the fall over the second session would be the result of the joint effect of repetition and feedback. Therefore, we expected PR to drop more intensely over the second session than over the first one.

Hypothesis 3. Differences by group: The comparisons among groups should reflect the effect of having used different types of outcomes in each group. In addition, the differences between Group 1 and the other two groups in the second session should be also presumably explained by the use of monetary incentives in Group 1. Taking this into account, we assumed that subjects in Group 2 would be more risk averse than those in Group 3, since the failure outcome in the former group (i.e. 'death') is more undesirable than the failure outcome in the latter group (i.e. 'days with back pain'). On this basis, we hypothesised that, as compared with respondents in Group 3, those in Group 2 would be more likely: a) to choose P over $; and b) to value each lottery at less than its expected value and, in consequence, to provide, for a same lottery, lower CE values than those reported by individuals in Group 3.[17] As a result, a priori, the difference between the two groups in terms of PR is indeterminate. We formulated no hypothesis regarding Group 1 in comparison with the other two groups because, as noted in the introduction, the use of monetary incentives seems to have a weak effect on PR (Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Camerer and Hogarth, 1999).

Hypothesis 4. Differences depending on the value offered as starting point in the CE questions: We conjectured that the largest proportion of PR would be obtained with the 'high' starting point. The reasoning is that the potential anchor value was larger with this format than with the other two (i.e. 'low' and 'open' starting points) and, therefore, we expected the $-bets to be valued more highly with the high starting point. Following the same argument, we hypothesised that the lowest rate of PR would be found with the 'low' starting point, since the potential anchor value was lower with this format than with the other two. Consequently, the proportion of PR with the open starting point would be comprised between the rates of PR obtained with the other two formats.

Hypothesis 5. Differences by gender: We presumed that women would be more risk averse than men, as supported by empirical evidence (Schubert et al., 1999; Hartog et al., 2002; Eckel and Grossman, 2008). Following the same rationale as in hypothesis 3, we conjectured

[17] We made this hypothesis because, according to the expected utility theory, given a bet (x, p; y) –where 'x' and 'y' are the success and failure outcomes of the lottery, respectively, and 'p' is the probability of success)–, whose expected value is given by px + (1-p)y, the decision maker is considered to be risk averse if the CE value that he/she places on the lottery is less than the expected value for that bet (Keeney and Raiffa, 1993).

that women would be more prone than men: a) to choose P over \$; and b) to value the gambles below their respective expected values and, therefore, to state, for a given bet, lower CE values than those provided by men. Thus, a priori, it is difficult to ascertain which of the two groups will have a higher rate of PR.

Hypotheses 1 and 2 were tested using the McNemar's test, because both of them entailed within-group comparisons. More specifically, in the case of hypothesis 2, the effect of repetition was tested making pairwise comparisons between the rounds of the first session, whereas the joint effect of repetition and feedback was tested making pairwise comparisons between the rounds of the second session. The remaining hypotheses (3-5) involved between-group comparisons, which were performed using the Fisher's exact test.

4.3. RESULTS

4.3.1. Background statistics

Table 4.4 presents the descriptive statistics corresponding to the elicited values for each bet. In this and all the subsequent tables, the rounds are labelled R_i-S_j, where 'i' stands for the number of round in each session (i = 1, 2, 3), and 'j' indicates the number of session (j = 1, 2). The rounds are numbered taking into account the order in which the same valuation (or the same choice) was made in each session. The first two numeric columns (labelled 'S1' and 'S2') display the summary rates for the first and second sessions, respectively.

Although the two paired gambles were matched or almost matched in their expected values, the distributions of the CE(P) and CE(\$) values were significantly different ($p = 0.000$ in all rounds). In particular, the mean and median CE(\$) values were larger than the mean and median CE(P) values. Moreover, the standard deviations and the interquartile ranges of the CE(\$) values were much larger than those of the CE(P) values. As a result, the CE(P) values were more concentrated around the expected values than the CE(\$) values, which could be explained because the range of possible values was more ample for the \$-bets (from 0 to 27 for lottery B and from 0 to 21 for lottery D) than for the P-bets (from 0 to 8 and from 0 to 5 for lotteries A and C, respectively). Nonetheless, the CE(\$) and CE(P) values tended to converge over the course of each session, due to the progressive decline of the former and to the flat trajectory of the latter. As can be seen in Table 4.4, the convergence between the mean and median CE(P) and CE($) values was stronger with the pair (AB) than with (CD).

For further details, Table 4.5 shows the proportions of respondents whose stated CE values were greater than, equal to or lower than the expected values of the gambles. The data are presented for each lottery and also for the average of the two P-bets and of the two $-bets. Most respondents (on average, 80.3% of subject in the first session and 72.1% in the second one) valued the P-bets at less than their expected values. On the contrary, the reverse pattern was observed with the $-bets. To be exact, 79.1% and 66.9% of respondents valued the $-bets above their respective expected values in sessions 1 and 2, respectively. As noted when formulating hypotheses 3 and 5, under expected utility, when an individual values a lottery at less than its expected value, he/she is regarded as ‘risk averse’. According to this, from Table 4.5 we can conclude that, in general, respondents displayed much stronger risk aversion when they valued a P-bet than when the lottery object of valuation was a $-bet.

Table 4.6 compares the elicited values for the two paired bets, indicating the proportions of subjects who reported higher CE values for the P-bet than for the $-bet (P > $), lower CE values for the $-bet than for the P-bet (P < $) or exactly the same value for the two paired bets (P = $). The figures are provided for each pair as well as for the average of the two pairs. The vast majority of respondents (on average, around 90% in the first session and 80% in the second one) attached a higher value to $ than to P. This pattern was more pronounced with (CD) than with (AB). On the contrary, 7% and 15.5% of subjects, on average, valued P more highly than $ in sessions 1 and 2, respectively. Approximately, only 4% and 6% of respondents equated the CE values of the two paired bets in sessions 1 and 2, respectively. As can be seen in Table 4.7, the response profile that emerges when comparing the CE(P) and CE($) values remained quite stable throughout the experiment, since only a small fraction of respondents switched their valuation behaviour. In general, switches from CE($) > CE(P) to CE(P) > CE($) were more frequent than the reverse direction switches, especially in the second session.

Table 4.4. CE values, summary statistics

		S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
A	Mean	6.3	6.4	6.3	6.2	6.2	6.3	6.5	6.5
	SD	1.5	1.5	1.4	1.5	1.5	1.5	1.5	1.5
	Median	7	7	7	7	7	7	7	7
	IQR	6-7.5	6-7.5	6-7.5	5.5-7.5	5.5-7.5	5.5-7.5	6-7.5	6-7.5

Table 4.4 (continued)

		S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
B	Mean	11.1	9.7	11.3	11	11.1	10.4	9.7	9.1
	SD	4.6	4.9	4.6	4.6	4.6	4.8	4.9	4.8
	Median	10	9	10	10	10	9.5	9	8
	IQR	8-14	6.5-12	8-15	8-13.5	8-14	7-12	6.5-12	6-11.5
C	Mean	3.3	3.5	3.4	3.3	3.3	3.4	3.5	3.5
	SD	0.86	0.93	0.87	0.86	0.86	0.89	0.90	0.99
	Median	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
	IQR	3-4	3-4	3-4	3-4	3-4	3-4	3-4	3-4
D	Mean	8.3	7.1	8.3	8.3	8.3	7.8	6.8	6.6
	SD	3.9	3.8	3.9	3.8	3.9	3.7	3.8	3.9
	Median	8	6.5	8	8	7.5	7	6	6
	IQR	5-10	4.5-9	5-10	5.5-10	5.5-10	5-10	4.5-8	4-8

N = 957 observations in each session for each lottery (319 observations per round).
SD: Standard deviation. IQR: Interquartile range.

Table 4.5. Comparison between CE and expected values

%		S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
(AB)	P > EV	3.9	10.7	4.7	3.8	3.1	5.0	12.5	14.7
	P = EV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	P < EV	96.1	89.3	95.3	96.2	96.9	95.0	87.5	85.3
	$ > EV	71.3	55.9	73.4	69.9	70.5	64.3	54.2	49.2
	$ = EV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$ < EV	28.7	44.1	26.6	30.1	29.5	35.7	45.8	50.8
(CD)	P > EV	10.1	18.3	11.3	9.7	9.4	14.4	19.7	20.7
	P = EV	25.5	26.9	24.5	27.3	24.8	27.6	27.0	26.0
	P < EV	64.4	54.9	64.3	63.0	65.8	58.0	53.3	53.3
	$ > EV	86.9	77.8	85.6	87.1	88.1	88.7	75.2	69.6
	$ = EV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$ < EV	13.1	22.2	14.4	12.9	11.9	11.3	24.8	30.4
Average	P > EV	7.0	14.5	8.0	6.7	6.3	9.7	16.1	17.7
	P = EV	12.7	13.4	12.2	13.7	12.3	13.8	13.5	13
	P < EV	80.3	72.1	79.8	79.6	81.4	76.5	70.4	69.3
	$ > EV	79.1	66.9	79.5	78.5	79.3	76.5	64.7	59.4
	$ = EV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$ < EV	20.9	33.1	20.5	21.5	20.7	23.5	35.3	40.6

(A, B) and (C, D): *N* = 957 observations in each session (319 observations per round).
Average: *N* = 1,914 observations in each session (638 observations per round).
The expected value (EV) of lotteries A, B, C and D is 7.6, 8.1, 4.0 and 4.2, respectively.

Table 4.6. Comparison between CE(P) and CE($) values

%		S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
(AB)	P > $	9.7	21.4	10.3	10.7	8.2	14.7	22.3	27.3
	P = $	3.7	6.6	3.8	2.2	5.0	3.4	6.6	9.7
	P < $	86.6	72.0	85.9	87.1	86.8	81.8	71.2	63.0
(CD)	P > $	4.2	9.5	4.7	4.7	3.1	4.7	11.9	11.9
	P = $	4.6	6.2	5.3	3.1	5.3	4.1	5.6	8.8
	P < $	91.2	84.3	90.0	92.2	91.5	91.2	82.4	79.3
Average	P > $	7.0	15.5	7.5	7.7	5.7	9.7	17.1	19.6
	P = $	4.2	6.4	4.6	2.7	5.2	3.8	6.1	9.3
	P < $	88.9	78.2	88.0	89.7	89.2	86.5	76.8	71.2

(A, B) and (C, D): N = 957 observations in each session (319 observations per round).
Average: N = 1,914 observations in each session (638 observations per round).

Table 4.7. Switch of CE values

%		Switch in:				
		R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
(AB)	From P > $ to $ > P	5.3	3.8	3.1	2.2	3.4
	From $ > P to P > $	4.7	1.9	8.8	9.1	6.6
	From P > $ to P = $	0.6	0.9	1.3	0.9	1.6
	From $ > P to P = $	1.6	3.1	1.6	5.0	6.6
	From P = $ to P > $	1.6	0.3	1.9	1.6	3.4
	From P = $ to $ > P	2.2	0.9	2.2	1.3	1.6
(CD)	From P > $ to $ > P	2.2	1.6	1.6	1.3	4.7
	From $ > P to P > $	4.7	1.9	8.8	9.1	6.6
	From P > $ to P = $	0.6	0.9	0.9	0.9	1.6
	From $ > P to P = $	1.6	3.1	1.6	5.0	6.6
	From P = $ to P > $	1.6	0.3	1.9	1.6	3.4
	From P = $ to $ > P	2.2	0.9	2.2	1.3	1.6
Average	From P > $ to $ > P	3.8	2.7	2.4	1.7	4.1
	From $ > P to P > $	4.7	1.9	8.8	9.1	6.6
	From P > $ to P = $	0.6	0.9	1.1	0.9	1.6
	From $ > P to P = $	1.6	3.1	1.6	5.0	6.6
	From P = $ to P > $	1.6	0.3	1.9	1.6	3.4
	From P = $ to $ > P	2.2	0.9	2.2	1.3	1.6

N = 319 observations in each round.
The percentages displayed in the 'R1-S2' column are the result of comparing the first round of valuations in session 2 with the third round of valuations in session 1. The figures in the rest of columns were calculated by comparing the second and subsequent rounds of valuations with the previous one.

Table 4.8. Choices

%		S1	S2	R1-S1	R2-S1	R3-S1
(AB)	P	58.4	72.1	58.3	59.2	57.7
	$	41.6	27.9	41.7	40.8	42.3
(CD)	P	58.4	68.0	57.4	61.4	56.4
	$	41.6	32.0	42.6	38.6	43.6
Average	P	58.4	70.0	57.9	60.3	57.0
	$	41.6	30.0	42.1	39.7	43.0

(A, B) and (C, D): *N* = 957 observations in the first session (319 observations per round); 319 observations in the second session.
Average: *N* = 1,914 observations in the first session (638 observations per round); 638 observations in the second session.
The proportions shown in column 'S2' correspond to those for the single choice of the second session.

Table 4.9. Switch of choices

%		Switch in:		
		R2-S1	R3-S1	S2
(AB)	From P to $	6.9	4.7	3.1
	From $ to P	7.8	5.3	11.6
(CD)	From P to $	6.0	5.0	5.6
	From $ to P	10.0	0.6	13.2
Average	From P to $	6.5	4.9	4.4
	From $ to P	8.9	3.0	12.4

N = 319 observations in each round.
The percentages in the first two columns were calculated by comparing the second and subsequent choices with the previous one. The figures in the last column are the result of comparing the choice of the second session with the third choice of the first session.

Turning to the choices, Table 4.8 details the percentages of subjects who chose P and $ in each round. In all choices, more than 50% of respondents favoured P over $ and the choice behaviour was quite similar with the two pairs of gambles. It is noteworthy that the preference for the P-bet was stronger in the choice of the second session than in any of the three choices of the first session. In addition, Table 4.9 shows that, at the individual level, the preference for P over $ generally persisted throughout the experiment, because only a small proportion of respondents who chose P switched their choice. Besides, in general, the number of subjects who changed their choice from $ to P was higher than the number of subjects who did the opposite. For instance, on average, 12.4% of respondents who chose $ in the third choice of the first session selected P in the choice of the following session, whereas only 4.4% of subjects who chose P in the third choice of the first session changed their preferred alternative in the choice of the next session.

4.3.2. Comparison between standard preference reversal and non-standard preference reversal at the aggregate level

Table 4.10 combines the information previously presented in Tables 4.6 and 4.8 and reports the proportions of the possible types of response obtained in each round. A 'response' refers to the result of comparing the CE values for P and $ with the choice between both bets. In the second session, since there were three valuations for each of the four gambles (A, B, C and D), but only one choice with each pair (A *vs.* B and C *vs.* D), rounds 1, 2 and 3 are the result of comparing the single choice with the first, second and third valuations for P and $, respectively. The responses are classified into four categories, which are labelled 'SPR', 'NSPR', 'Consistent' and 'Equal'. Respondents incurred in SPR if they chose P but valued $ more highly. The reverse pattern (i.e. if $ was chosen and the value assigned to P was higher) is identified as NSPR. Subjects were consistent when the chosen lottery was valued more highly than its paired bet. Finally, the label 'Equal' indicates that the same value was placed on the two paired lotteries. Although there was no indifferent option in the choices, we assume that respondents who attached the same value to P and $ were indifferent between the two alternatives.[18]

Consistently with the figures provided in the preceding tables, Table 4.10 reveals that most respondents incurred in SPR (i.e. they chose P but ascribed a higher value to $ than to P). More specifically, the average rate of SPR was 50% in the first session and 53% in the second one. By contrast, almost no subject displayed the opposite pattern: on average, 1.5% and 2.9% of subjects committed NSPR in sessions 1 and 2, respectively. As a result, SPR was significantly more frequent than NSPR throughout the experiment (p = 0.000 in all rounds), thus confirming our first hypothesis. The same strong asymmetry between the two types of PR was found in all groups, as it will be shown below. Moreover, considering the two pairs of lotteries as a whole, the number of respondents who incurred in SPR was significantly higher than the

[18] Notwithstanding, we are aware that some authors that have allowed for indifference in the choices (e.g. Oliver 2006, 2013a, 2013b) interpret that subjects who choose P and ascribe the same value to P as to $ incur in a 'weak predicted reversal', and that those who choose $ and put the same value on the two bets commit a 'weak unpredicted reversal'.

number of consistent respondents (p = 0.014 in the first session and p = 0.000 in the second one).

Comparing the results obtained with the two pairs, SPR was more frequent with (CD) than with (AB), while NSPR was less frequent with the former pair than with the latter. Consequently, the asymmetric pattern of PR was more noticeable with (CD) than with (AB). Between 30% and 43% of subjects (depending on the round) committed SPR with both pairs in the same round, whereas no more than 1.3% of respondents exhibited NSPR with the two pairs in a same round. Furthermore, in all rounds the number of individuals who incurred in SPR with (CD) (more than a half of the sample) was higher than the number of consistent respondents. The proportion of SPR with (AB) was also high, but to a lesser extent (it only surpassed 50% in round 1 of the second session), and in some rounds the number of consistent respondents with this pair was higher than the number of those who showed SPR.

Apart from being very frequent, the presence of SPR was also quite persistent. In particular, 22.9% of the sample displayed SPR in each of the three rounds of the first session and with both pairs of lotteries, whilst 24.8% presented this pattern in the three rounds of the second session. The persistence of SPR suggests that it is a systematic anomaly rather than a random one. Conversely, in both sessions there was no subject who incurred in NSPR in the three rounds of the same session and with the two pairs of gambles.

Table 4.10. Classification of responses at the aggregate level

%		S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
(AB)	SPR	47.3	49.1	48.0	47.7	46.4	56.7	48.3	42.3
	NSPR	1.7	3.6	2.8	1.3	0.9	2.5	3.1	5.0
	Consistent	47.3	40.8	45.5	48.9	47.7	37.3	42.0	42.9
	Equal	3.7	6.6	3.8	2.2	5.0	3.4	6.6	9.7
(CD)	SPR	52.6	56.8	51.1	55.5	51.1	61.8	54.9	53.9
	NSPR	1.4	2.3	1.9	0.9	1.3	1.3	2.5	3.1
	Consistent	41.5	34.7	41.7	40.4	42.3	32.9	37.0	34.2
	Equal	4.6	6.2	5.3	3.1	5.3	4.1	5.6	8.8
Average	SPR	50.0	53.0	49.5	51.6	48.8	59.3	51.6	48.1
	NSPR	1.5	2.9	2.4	1.1	1.1	1.9	2.8	4.1
	Consistent	44.4	37.7	43.6	44.7	45.0	35.1	39.5	38.6
	Equal	4.1	6.4	4.6	2.7	5.2	3.8	6.1	9.3

(A, B) and (C, D): N = 957 observations in each session (319 observations per round).
Average: N = 1,914 observations in each session (638 observations per round).

4.3.3. The effects of repetition and feedback at the aggregate level

Table 4.11 summarises the tests for the effects of repetition and feedback (the significant differences are highlighted in bold, as well as in the subsequent tables). The repetition of valuation and choices over the first session had a modest effect on the pattern observed in these tasks. Consequently, in general, neither the changes between consecutive rounds nor the changes between rounds 1 and 3 of the first session were statistically significant. As can be seen in Table 4.10, between these two rounds, the overall rate of PR (SPR + NSPR) decreased from 51.9% to 49.9%, on average, due to both a slight fall in SPR (from 49.5% to 48.8%) and in NSPR (from 2.4% to 1.1%). In consequence, consistency improved only moderately (from 43.6% to 45% of responses). As a result of the trend experienced by SPR and NSPR over the first session, the asymmetry of PR increased a little. In summary, the repetition of tasks over the first session did not have a significant effect on any of the two types of PR. This means that respondents' stated preferences were rather insensitive to the mere repetition of tasks.

The comparison between the two pairs shows that PR was a bit less frequent in round 3 than in round 1of the first session in both cases, but this variation was more marked with the pair (AB) (47.3% *vs.* 50.8%, respectively) than with (CD) (52.4% *vs.* 53%, respectively). Indeed, SPR with the latter pair was equally frequent in both rounds (51.1%). Therefore, the improvement in consistency was larger with (AB) than with (CD). Another difference between the two pairs was observed in round 2, in which SPR rose significantly with (CD) (from 51.1% to 55.5%; $p = 0.087$), while it declined a little with (AB) (from 48% to 47.6%). In both cases the frequency of NSPR was more reduced in round 3 than in round 1 of the first session, but the difference between these two rounds was not significant with any of the two pairs.

As happened in the first session, the overall rate of PR also decayed between rounds 1 and 3 of the next session (from 61.2% to 52.2%, on average) but in this case the difference was significant ($p = 0.000$). Another remarkable difference between the two sessions is that, whereas SPR and NSPR subsided over the first session (although not significantly), the two types of PR moved in opposite directions over the second session. Specifically, in this latter session, the proportion of SPR experienced a downward trend (from 59.3% to 48.1%; $p = 0.000$), but the proportion of NSPR increased (from 1.9% to 4.1%; $p = 0.003$). Therefore, the combination of repetition and feedback had a larger effect on PR than the repetition of tasks alone, thus confirming hypothesis 2. As a result of the decline in SPR and the rise in NSPR, the asymmetric pattern of PR softened over the second session, although it was still noteworthy in the last round of the experiment. Another consequences was that consistency improved significantly between rounds 1 and 3 of the second session (from 35.1% to 38.6%; $p = 0.019$). Thus, the combination of repetition of tasks and feedback made respondents be more consistent, unlike repetition alone. However, despite this improvement, at the end of the experiment, SPR was still significantly more frequent than consistency (48.1% *vs.*38.6% of responses, on average; p = 0.011). On closer inspection, the drop in SPR (between rounds 1 and 3 of the second session) was significant with both pairs of gambles, but the upward trends in NSPR and in consistency were significant only with (AB).

Table 4.11. Effects of repetition and feedback at the aggregate level

P-values		Repetition (first session)			Repetition+ feedback (second session)		
		R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3
(AB)	SPR	1.000	0.720	0.653	**0.000**	**0.008**	**0.000**
	NSPR	0.227	1.000	0.146	0.688	**0.070**	**0.022**
	Consistent	0.272	0.708	0.505	**0.044**	0.780	**0.039**
	Equal	0.359	**0.049**	0.541	**0.087**	0.164	**0.001**
(CD)	SPR	**0.087**	**0.039**	1.000	**0.000**	0.761	**0.000**
	NSPR	0.453	1.000	0.727	0.289	0.754	0.109
	Consistent	0.712	0.441	0.902	0.047	0.222	0.644
	Equal	0.189	0.092	1.000	0.473	0.133	**0.017**
Average	SPR	0.291	0.105	0.733	**0.000**	**0.026**	**0.000**
	NSPR	**0.096**	1.000	0.115	0.180	**0.096**	**0.003**
	Consistent	0.623	0.923	0.510	**0.003**	0.606	**0.045**
	Equal	**0.081**	**0.005**	0.659	**0.067**	**0.031**	**0.000**

Exact p-values from the McNemar's test.

In short, whilst the mere repetition of tasks (over the first session) did not have a significant influence on respondents' stated preferences, the combination of repetition and feedback (over the second session) reduced the frequency of SPR significantly but, in turn, the frequency of NSPR also increased significantly. Overall, the rate of consistent responses was significantly higher at the end of the second session than at the beginning of the same session, but it was still below 50% and lower than the proportion of SPR.

The following subsections provide separate analyses by groups, starting points presented in the valuation tasks and gender. In all cases, the proportions of the different types of responses were compared across groups, whereas the effects of repetition and feedback were tested independently for each group.

4.3.4. Analysis by groups

This subsection compares the results obtained in Groups 1, 2 and 3. In the first session, the comparisons among the three groups revealed the effect of the kind of outcomes used (money, years of life and days without back pain, respectively). In the second session, the same comparisons reflected the influence of the type of outcomes as well as the effect of the provision of monetary incentives in Group 1.

Table 4.12 shows that in the two sessions the overall rates of PR (SPR + NSPR) were very large in each group, ranging from approximately 45-60% in both Groups 1 and 3 to 54-65% in Group 2, depending on the round. The typical asymmetry of PR was observed in the three groups, with the proportion of SPR surpassing by far that of NSPR. This disparity was more salient in Group 2 than in the other two groups. Overall, the highest and the lowest percentages of SPR were found in the groups where the success outcomes of the lotteries were expressed as years of life (Group 2) and money (Group 1), respectively. The differences between these two groups in terms of the rate of SPR were significant in the two sessions, to a greater extent in the first session than in the second one. More specifically, on average, 46.5% and 58.4% of respondents belonging to Group 1 and to Group 2 incurred in SPR in the first session, respectively (p = 0.000), whereas the same rates increased up to 51.3% and 56.5% in the second session (p = 0.063). In consequence, the proportion of consistent respondents was higher in Group 1 than in Group 2 (45.7% *vs*. 39% in the first session and 37.7% *vs*. 35.4% in the second one), although the difference was significant only in the first session (p = 0.017). However, it must be noted that Group 1 had the highest rates of NSPR in the two sessions (1.9% and 3.7% in sessions 1 and 2, respectively), although the differences between any two of the three groups with regard to the proportion of NSPR were not significant. Group 3 exhibited a similar response profile to that of Group 1 in both sessions (although the rates of SPR were higher in the former group than in the latter), whereas the differences observed between Group 3 and Group 2 (in relation to the proportions of SPR and consistent responses) were significant in the first session.

The higher rates of SPR in Group 2 as compared with Group 3 could be explained because subjects who are presented with lotteries that involve a certain risk of death are possibly more risk averse than those who deal with lotteries that do not entail that risk. This assumption is supported by the results that we observed in the choice tasks. As can be seen in Table 4.13, in all choices, the proportion of subjects in Group 2 who preferred P over $ was significantly higher than the proportion of respondents in Group 3 who chose P (on average, 62.7% *vs*. 54.7% in the first session; p = 0.013; 76% *vs*. 67.9% in the second one; p = 0.006), confirming what we conjectured in hypothesis 3. The same conclusion is reached if Group 2 is compared with Group 1. Consistently, the proportion of subjects who preferred P over $ in all choices of the experiment was higher in Group 2 (39.2%) than in both Group 1 (29.6%) and Group 3 (25.9%).

However, our results do not support the assumption that subjects in Group 2 would be more likely than those in Group 3 to value the lotteries at less than their expected values. As a matter of fact, in the first session, on average, 78.5% of subjects in Group 2 placed on the P-bets CE values that were lower than the expected values of that lotteries, against 87% of respondents in Group 3 who had this behaviour ($p = 0.000$). However, in the second session, on average, the proportion of subjects who valued the $-bets below their expected values was greater in Group 2 than in Group 3 (34.4% and 27.2%, respectively; $p = 0.017$). Furthermore, the CE values obtained for a given bet were similar in both groups.

In regards to the effect of the use of monetary incentives in Group 1 in the second session, our results seem to suggest that the incentives had a rather limited effect. For instance, although Group 1 had the lowest rates of SPR in the two sessions, the differences between this group and the other two in relation to the proportion of SPR were smaller in the second session than in the first one. On the other hand, the difference between Group 1 and Group 2 in terms of the consistency rate (which was higher in the first one in both sessions) was also more reduced in session 2 than in session 1. In addition, the upward trend in NSPR over the second session was more marked in Group 1 than in the other two groups (especially in comparison with Group 3).

Table 4.14 presents, for each group, the *p*-values resulting from the within-subject comparisons between rounds. As far as the first session is concerned, the most notable result is that SPR only subsided significantly in Group 1, while it experienced an upward trend in the other two groups. To be exact, as shown in Table 4.12, between rounds 1 and 3 of that session, the rate of SPR in Group 1 fell from 49.4% to 43.1% ($p = 0.031$). Conversely, the same rate rose slightly in the other two groups –from 55.7% to 58.9% in Group 2 and from 43.8% to 50% in Group 3. In the three groups NSPR declined over the first session (but not significantly). Indeed, no respondent in Group 2 showed NSPR in round 3. As a result of the described trends in SPR and NSPR over the first session, consistency improved significantly in Group 1 (from 41.8% to 47.8%; $p = 0.048$), and the asymmetric pattern of PR was slightly attenuated. By contrast, the proportion of consistent responses diminished a little in Group 2 (from 40.5% to 39.2%) and in Group 3 (from 50% to 45.1%).

Turning to the second session, it is worth pointing out that in the three groups SPR was more frequent in round 1 of that session than in any other round of the experiment. It is especially striking that 63.9% of respondents in Group 2 committed SPR in this round. In the three groups the combination of repetition and feedback over the second session gave rise to a drop in the proportion of SPR and, consequently, to an increase in the rate of consistent responses, as expected. As can be seen in Table 4.14, the erosion of SPR was significant in the three groups. More specifically, between round 1 and round 3 of the second session, the rate of SPR fell from 57.6% to 45.3% in Group 1 (p = 0.000), from 63.9% to 51.3% in Group 2 (p = 0.000) and from 58% to 50.6% in Group 3 (p = 0.017). However, in the three cases the improvement in terms of consistency was limited by the upward trend in NSPR, although this rise was significant only in Group 1, where the rate of NSPR increased from 2.2% to 5.4% (p = 0.013). On average, at the end of the experiment, respondents in all groups were still more susceptible to exhibit SPR than a consistent pattern and still fewer than 40% of subjects in each group were consistent.

Table 4.12. Classification of response by groups

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
SPR								
G1	46.5	51.3	49.4	47.2	43.1	57.6	50.9	45.3
G2	58.4	56.5	55.7	60.8	58.9	63.9	54.4	51.3
G3	48.4	52.9	43.8	51.2	50.0	58.0	50.0	50.6
G1 *vs*. G2	**0.000**	**0.063**						
G1 *vs*. G3	0.539	0.577						
G2 *vs*. G3	**0.002**	0.271						
NSPR								
G1	1.9	3.7	2.8	1.3	1.6	2.2	3.5	5.4
G2	1.1	2.3	1.9	1.3	0.0	1.3	3.2	2.5
G3	1.2	2.1	1.9	0.6	1.2	1.9	1.2	3.1
G1 *vs*. G2	0.273	0.204						
G1 *vs*. G3	0.514	0.110						
G2 *vs*. G3	1.000	0.828						
Consistent								
G1	45.7	37.7	41.8	47.5	47.8	36.5	37.7	39.0
G2	39.0	35.4	40.5	37.3	39.2	30.4	38.6	37.3
G3	47.1	39.9	50.0	46.3	45.1	37.0	43.8	38.9
G1 *vs*. G2	**0.017**	0.415						
G1 *vs*. G3	0.615	0.423						
G2 *vs*. G3	**0.013**	0.162						

Table 4.12 (continued)

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
Equal								
G1	5.9	7.3	6.0	4.1	7.6	3.8	7.9	10.4
G2	1.5	5.7	1.9	0.6	1.9	4.4	3.8	8.9
G3	3.3	5.1	4.3	1.9	3.7	3.1	4.9	7.4
G1 *vs*. G2	**0.000**	0.266						
G1 *vs*. G3	**0.040**	0.117						
G2 *vs*. G3	**0.090**	0.776						

Average results for pairs (AB) and (CD). *N* = 954, 474 and 486 observations in each session for Group 1 (G1), Group 2 (G2) and Group 3 (G3), respectively (318, 158 and 162 observations per round).
The figures provided for the comparisons (e.g. G1 *vs.* G2) are *p*-values from the Fisher's exact test.

Table 4.13. Proportions of respondents (per group) who exhibited risk aversion in the choice and valuation tasks[a]

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
Choose P								
G1	58.2	68.2	60.1	60.1	54.5	68.2		
G2	62.7	76.0	62.1	64.0	62.0	76.0		
G3	54.7	67.9	49.4	57.4	57.4	67.9		
G1 *vs*. G2	0.109	**0.002**						
G1 *vs*. G3	0.216	0.905						
G2 *vs*. G3	**0.013**	**0.006**						
P < EV								
G1	77.7	69.8	77.0	77.4	78.6	73.0	70.1	66.4
G2	78.5	74.9	77.2	80.4	77.9	78.5	71.5	74.7
G3	87.0	73.7	87.7	83.3	90.1	81.5	69.8	69.8
G1 *vs*. G2	0.786	**0.047**						
G1 *vs*. G3	**0.000**	0.140						
G2 *vs*. G3	**0.000**	0.712						
$ < EV								
G1	25.7	35.5	24.2	27.0	25.8	26.7	36.2	43.7
G2	15.6	34.4	13.9	17.7	15.2	22.2	38.6	42.4
G3	16.7	27.2	19.8	14.2	16.1	18.5	30.3	32.7
G1 *vs*. G2	**0.000**	0.681						
G1 *vs*. G3	**0.000**	**0.001**						
G2 *vs*. G3	0.662	**0.017**						

Average results for pairs (AB) and (CD). *N* = 318, 158 and 162 observations per round for Group 1 (G1), Group 2 (G2) and Group 3 (G3), respectively.
The figures provided for the comparisons (e.g. G1 *vs.* G2) are *p*-values from the Fisher's exact test.
[a] We considered that those who chose P were more risk averse than those who chose $, as well as those who valued the gambles at less than their expected values as compared with those who did the opposite.

Table 4.14. Effects of repetition and feedback by groups

P-values	Repetition (first session)			Repetition+ feedback (second session)		
	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3
SPR						
G1	0.494	0.111	**0.031**	**0.002**	**0.008**	**0.000**
G2	0.152	0.678	0.442	**0.003**	0.442	**0.000**
G3	**0.036**	0.856	0.110	**0.002**	1.000	**0.017**
NSPR						
G1	0.180	1.000	0.388	0.289	0.109	**0.013**
G2	1.000	0.500	0.250	0.250	1.000	0.625
G3	0.625	1.000	1.000	1.000	0.375	0.500
Consistent						
G1	**0.063**	1.000	**0.048**	0.636	0.636	0.350
G2	0.458	0.678	0.856	**0.024**	0.845	**0.071**
G3	0.405	0.851	0.230	**0.013**	0.185	0.678
Equal						
G1	0.327	**0.027**	0.473	**0.041**	0.280	**0.001**
G2	0.625	0.500	1.000	1.000	**0.096**	0.144
G3	0.344	0.453	1.000	0.581	0.481	0.144

Exact *p*-values from the McNemar's test.

4.3.5. Analysis by starting point presented in the valuation tasks

As mentioned in the section of Methods, each group was further subdivided by taking into account the starting point value presented in the CE questions: low, high and open starting points. Table 4.15 classifies the responses obtained depending on the starting point format used in these tasks.

First, against hypothesis 4, the reported CE values seem to have been rather insensitive to the value presented as starting point because, in general, the mean CE values obtained with the three formats did not differ significantly. In terms of SPR, the comparison among the three groups revealed a differential pattern in each session. In the first one, the average rate of SPR ranged from 44.3% with the low starting point to 55.4% with the open starting point (p = 0.000). By contrast, in the second session, on average, the lowest and the highest proportions of SPR were obtained with the high and low starting points, respectively (48.4% *vs.* 56.9%; p = 0.002). Therefore, in the latter session, the observed pattern of SPR depending on the starting point provided in the CE questions was just the opposite of that we expected to find. On the other hand, the highest percentages of NSPR were obtained with the low starting point (on average, 2.3% and 4.3% in sessions 1 and 2, respectively), whereas the lowest rates were found with the high starting point (0.8% in the first session) and with the open starting point (1.6% in the second one). As a result, in both sessions the high starting point led to higher average rates of consistent responses than the open starting point (44.9% *vs*. 39.5%; $p = 0.053$ in the first session, and 43.3% *vs*. 37.7%; $p = 0.045$ in the second one). However, there was no clear pattern when comparing the high and low starting points: whilst in the first session the average consistency rate was higher with the low than with the high starting point, the opposite was true in the second session.

Turning to the trends in the different types of responses, Table 4.15 and Table 4.16 show that there were little changes over the first session in the three groups. Indeed, on average, there were no significant variations between rounds 1 and 3 of this session in any of the three cases. In general, both SPR and NSPR decreased over the first session regardless of the starting point given in the valuation tasks. The only exception was observed with the open starting point, which resulted in a slightly higher rate of SPR in round 3 (54.3%) than in round 1 (52.4%) of the first session.

On the contrary, significant changes were found over the second session (with repetition and feedback). In particular, comparing rounds 1 and 3 of this session, the proportion of SPR subsided significantly in the three groups: from 63.4% to 51.4% with the low starting point; from 54.3% to 44.8% with the high starting point; and from 60% to 48.1% with the open starting point ($p = 0.000$ in the three cases). By contrast, the frequency of NSPR went upwards with the three groups, thus limiting the improvement in the consistency rate, which did not increase significantly in any of the three cases. The rise in NSPR between rounds 1 and 3 of the second session was significant with the low starting point (from 2.8% to 6.5%; $p = 0.022$) and with the high starting point (from 1.4% to 3.8%; $p = 0.063$), but not with the open starting point (from 1.4% to 1.9%; $p = 1.000$).

Table 4.15. Classification of responses by starting point format of the valuation tasks

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
SPR								
Low	44.3	56.9	44.9	44.0	44.0	63.4	56.0	51.4
High	50.3	48.4	51.4	51.4	48.1	54.3	46.2	44.8
Open	55.4	53.5	52.4	59.4	54.3	60.0	52.4	48.1
Low *vs*. high	**0.033**	**0.002**						
Low *vs*. open	**0.000**	0.217						
High *vs*. open	**0.081**	**0.081**						
NSPR								
Low	2.3	4.3	3.2	1.9	2.3	2.8	3.7	6.5
High	0.8	2.9	1.4	0.5	0.5	1.4	3.3	3.8
Open	1.3	1.6	2.4	0.9	0.5	1.4	1.4	1.9
Low *vs*. high	**0.026**	0.178						
Low *vs*. open	0.148	**0.005**						
High *vs*. open	0.579	0.130						
Consistent								
Low	48.8	32.3	47.2	50.9	48.2	30.6	34.3	31.9
High	44.9	43.3	43.8	44.3	46.7	39.5	46.7	43.8
Open	39.5	37.7	39.6	38.7	40.1	35.4	37.7	40.1
Low *vs*. high	0.179	**0.000**						
Low *vs*. open	**0.001**	**0.041**						
High *vs*. open	**0.053**	**0.045**						
Equal								
Low	4.5	6.5	4.6	3.2	5.6	3.2	6.0	10.2
High	4.0	5.4	3.3	3.8	4.8	4.8	3.8	7.6
Open	3.9	7.2	5.7	0.9	5.2	3.3	8.5	9.9
Low *vs*. high	0.679	0.478						
Low *vs*. open	0.678	0.659						
High *vs*. open	1.000	0.204						

Average results for pairs (AB) and (CD). N = 648, 630 and 636 observations in each session for the low, high and open formats, respectively (216, 210 and 212 observations per round).
The figures provided for the comparisons (e.g. low *vs*. high) are p-values from the Fisher's exact test.

Table 4.16. Effects of repetition and feedback by format of the valuation tasks

	Repetition (first session)			Repetition+ feedback (second session)		
P-values	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3
SPR						
Low	0.883	1.000	0.880	**0.005**	0.110	**0.000**
High	1.000	0.360	0.410	**0.002**	0.690	**0.000**
Open	**0.032**	**0.061**	0.636	**0.003**	0.163	**0.000**
NSPR						
Low	0.508	1.000	0.754	0.727	**0.070**	**0.022**
High	0.625	1.000	0.625	0.125	1.000	**0.063**
Open	0.375	1.000	0.219	1.000	1.000	1.000
Consistent						
Low	0.350	0.430	0.883	0.201	0.487	0.755
High	1.000	0.522	0.512	**0.008**	0.345	0.163
Open	0.885	0.701	1.000	0.442	0.487	0.133
Equal						
Low	0.581	0.180	0.791	0.263	0.176	**0.006**
High	1.000	0.754	0.581	0.804	**0.077**	0.286
Open	**0.013**	**0.012**	1.000	**0.035**	0.701	**0.004**

Exact *p*-values from the McNemar's test.

4.3.6. Analysis by gender

Table 4.17 details the classification of responses for both men and women. The most prominent result is that women showed SPR more frequently than men throughout the experiment. The difference was statistically significant in the two sessions, but it was more noticeable in the second one. On average, 41.4% of men and 55.9% of women committed SPR in the first session, while 42.9% of men and 60% of women displayed this pattern in the second one (p = 0.000 in both sessions). On the contrary, the proportions of NSPR were barely the same in both groups and in the two sessions: on average, 1.7% of men and 1.4% of women showed NSPR in the first session, while the same proportions rose to 2.8% and 3%, respectively, in the second one. Since SPR was far more frequent than NSPR in both groups, the asymmetric pattern of PR was stronger among women than among men, and the rates of consistent responses were significantly lower in the former group than in the latter (39.3% *vs.* 51.8% and 30.7% *vs.* 47.8% in sessions 1 and 2, respectively; p = 0.000 in both sessions). The divergence between the consistency rates of men and women was larger in the second session than in the first one.

The higher proportions of SPR in the female group than in male group are mainly explained because women revealed a stronger preference for P over $ in the choice tasks (i.e. women chose P at a higher rate than men) (see Table 4.18 for details). On average, approximately 51% of men and 63% of women chose P in the first session (p = 0.000). Similarly, in the single choice of the second session, 58.8% of men and 77.9% of women preferred P (p = 0.000). Likewise, the proportion of subjects who always chose P was significantly higher in the female group than in the male group, whereas the opposite occurred with those who chose $ in all cases. To be exact, 21.4% of men and 37.8% of women favoured P over $ in all choices of the experiment (p = 0.002). These results might suggest that women in this experiment were more risk averse than men when choosing, given that the P-bets offered a higher probability of success than the $-bets.

Nevertheless, the stronger risk aversion of women was not so clearly reflected in the valuation tasks. On the one hand, in both sessions, the proportions of respondents who valued the lotteries at less than their expected values were higher in the female group than in the male group, on average. However, the opposite happened with the $-bets in the first session, although in the second session almost the same proportions of men and women valued the $-bets below their expected values. In conclusion, the data in Table 4.18 suggests that the gender difference observed in terms of SPR was primarily driven by a greater propensity of women to choose P over $.

Although to a lesser extent, the higher rates of SPR in the female group than in the male group are also attributable to differences in the elicited values. In general, women tended to state lower values for the P-bets and higher values for the $-bets than those reported by men. Additionally, women's CE values had a higher dispersion than men's (i.e. the standard deviations and the interquartile ranges of the valuations were generally greater among women than among men).

Table 4.19 displays, independently for men and women, the p-values resulting from the tests for the effect of repetition and for the combined effect of repetition and feedback. In general, both men's and women's response behaviours were quite insensitive to the repetition of tasks over the first session. In the male group, on average, SPR was slightly more frequent in round 1 than in round 3 of that session, whereas the opposite happened in the female group. In both groups, NSPR declined between round 1 and round 3 of the first session, but not significantly. The rates of consistent men and women moved in opposite directions over the same session. Specifically, while the proportion of consistent men was lower in round 3 (51.5%) than in round 1 (54.2%), the rate of consistent women increased between the two rounds (from

36.2% to 40.4%). In any case, the variations in the consistency rates between rounds 1 and 3 of the first session were not significant in any of the two groups.

Moving to the figures pertaining to the second session, Table 4.20 reveals that the joint effect of repetition and feedback over this session had significant effects on both men's and women's stated preferences. In the two groups the decline in PR was more pronounced than in the first session, but this fall was entirely driven by the downward trend in SPR, inasmuch NSPR became more frequent over the second session. The variations in the frequencies of SPR and NSPR between rounds 1 and 3 of that session were significant in the two groups. Specifically, the percentage of SPR dropped from 48.5% to 36.6% in the male group and from 66.8% to 56.1% in the female group (p = 0.000 in both cases). Conversely, the proportion of NSPR rose from 1.2% to 3.8% among men (p = 0.039) and from 2.4% to 4.3% among women (p = 0.065). Although the rate of consistent respondents increased between rounds 1 and 3 in the two groups, the difference was significant only in the female group. To be exact, the rates of consistent men and women rose from 46.6% to 48.5% (p = 0.568) and from 27.1% to 31.7% (p = 0.040), respectively. Despite the greater improvement in the latter group, in the last round of the experiment there was still a significantly lower proportion of consistent women than of men (p = 0.000).

Table 4.17. Classification of responses by gender

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
SPR								
Men	41.4	42.9	39.3	45.0	39.7	48.5	43.5	36.6
Women	55.9	60.0	56.7	56.1	55.1	66.8	57.2	56.1
P-values	**0.000**	**0.000**						
NSPR								
Men	1.7	2.8	2.3	1.2	1.5	1.2	3.4	3.8
Women	1.4	3.0	2.3	1.1	0.8	2.4	2.4	4.3
P-values	0.706	0.891						
Consistent								
Men	51.8	47.8	54.2	49.6	51.5	46.6	48.5	48.5
Women	39.3	30.7	36.2	41.2	40.4	27.1	33.2	31.7
P-values	**0.000**	**0.000**						
Equal								
Men	5.2	6.5	4.2	4.2	7.3	3.8	4.6	11.1
Women	3.4	6.3	4.8	1.6	3.7	3.7	7.2	8.0
P-values	**0.048**	0.924						

Average results for pairs (AB) and (CD). N = 786 and 1,128 observations in each session for men and women, respectively (262 and 376 observations per round).
The p-values for the comparisons between men and women were derived from the Fisher's exact test.

Table 4.18. Proportions of men and women who exhibited risk aversion in the choice and valuation tasks[a]

%	S1	S2	R1-S1	R2-S1	R3-S1	R1-S2	R2-S2	R3-S2
Choose P								
Men	51.0	58.8	48.9	55.3	48.9	58.8		
Women	63.6	77.9	64.1	63.8	62.8	77.9		
P-values	**0.000**	**0.000**						
P < EV								
Men	76.1	67.8	77.5	74.4	76.3	71.8	67.2	64.5
Women	83.2	75.0	81.4	83.2	84.8	79.8	72.6	72.6
P-values	**0.000**	**0.001**						
$ < EV								
Men	23.7	32.1	21.4	24.4	25.2	23.3	33.6	39.3
Women	19.0	33.9	20.0	19.4	17.6	23.7	36.4	41.5
P-values	**0.014**	0.430						

Average results for pairs (AB) and (CD). N = 262 and 376 observations per round for men and women, respectively.
The *p*-values for the comparisons between men and women were derived from the Fisher's exact test.
[a] We considered that those individuals who chose P were more risk averse than those who chose $, as well as those who valued the gambles at less than their expected values.

Table 4.19. Effects of repetition and feedback by gender

P-values	Repetition (first session)			Repetition+ feedback (second session)		
	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3	R1 *vs.* R2	R2 *vs.* R3	R1 *vs.* R3
SPR						
Men	**0.036**	**0.049**	1.000	**0.019**	**0.003**	**0.000**
Women	0.913	0.712	0.586	**0.000**	0.689	**0.000**
NSPR						
Men	0.453	1.000	0.727	**0.070**	1.000	**0.039**
Women	0.227	1.000	0.146	1.000	**0.065**	**0.065**
Consistent						
Men	0.126	0.560	0.435	0.487	1.000	0.568
Women	**0.067**	0.795	0.109	**0.002**	0.500	**0.040**
Equal						
Men	1.000	**0.057**	0.115	0.815	**0.005**	**0.001**
Women	**0.081**	**0.005**	0.659	**0.060**	0.766	**0.017**

Exact *p*-values from the McNemar's test.

4.4. DISCUSSION

An extensive body of research has found the preference reversal (PR) phenomenon to be a robust and persistent violation of procedural invariance. This study was primarily motivated by evidence that PR is attenuated in some repeated experimental markets. In this respect, the main objective of this study was to test whether the presence of PR declines as individuals learn through the repetition of tasks and through the feedback on the consequences of their decisions. To that end, we conducted an experiment involving two sessions, so as to separate the effect of the repetition of tasks (first session) from the effect of repetition along with feedback (second session). Another important objective was to examine whether the use of health outcomes yields different results as those obtained with monetary outcomes. In addition, we analysed differences in the rate of PR depending on the starting point presented in the valuation questions, as well as gender differences in terms of PR.

We found the typical asymmetric pattern of PR in both sessions, with standard preference reversal (SPR) being much more frequent than non-standard preference reversal (NSPR). Albeit PR decayed with experience, the effect was significant only in the second session. This means that the learning effect which stemmed from the mere repetition of tasks was limited. By contrast, when respondents had the opportunity to learn from both repetition and feedback, the overall rate of PR (SPR + NSPR) fell substantially. Nonetheless, this improvement was only explained by the subsidence of SPR, given that the frequency of NSPR increased over the second session. But, despite these two opposite trends, SPR was still remarkably frequent at the end of the experiment (around 50% of responses at the aggregate level), whereas the rate of NSPR was always below 6%. As a result, the asymmetry between SPR and NSPR slightly increased over the first session, whilst it lessened over the second one.

To some extent, the results obtained in the second session resemble those reported by Braga et *al.* (2009). Like us, these researchers also found that the frequency of SPR subsided in a second-price Vickrey auction (SPA) with repetition, feedback and incentives, while the frequency of NSPR increased. But, unlike in our experiment, they observed that the two types of PR first converged when feedback was provided and then NSPR became even more frequent than SPR.

The drop in the frequency of SPR (especially over the second session) is consistent with the discovered preference hypothesis (Plott, 1996). According to this hypothesis, individuals have a set of stable underlying preferences, but, when they undertake unfamiliar tasks, they may not be clear about what they really prefer. However, individuals may discover their own preferences as they acquire experience with the tasks they are asked to make, for instance, if they perform them several times, receive feedback on the consequences of the decisions adopted in those tasks, as well as incentives to think carefully about those consequences. Conversely, the upward trend NSPR observed over the second session of our experiment is against the predictions of the aforementioned hypotheses. A different interpretation of the subsidence of SPR consists in supposing that subjects participants in our experiment did not have well-defined (or complete) preferences, but they were constructed depending on a number of factors, such as the elicitation tasks, the starting point presented in the valuation questions or the specific gambles being assessed.

To the best of our knowledge, our study has been the first to examine PR in relation to both monetary and health outcomes. Some prior studies have compared PR regarding money and non-monetary outcomes others than health. For example, Slovic *et al.* (1990) undertook an experiment involving six pairs of bets, half of which offered monetary payoffs and the other half offered different types of non-monetary outcomes, such as a one-week pass for the theatre or a dinner for two people. They found that the proportion of SPR was significantly lower with the non-monetary bets (24%) than with the monetary bets (41%), which might be ascribable to the compatibility between prices and monetary outcomes. Our results led to opposite conclusions, insofar, on average, the lowest and the highest rates of PR (SPR) were obtained in Group 1 ('money') and Group 2 ('years of life'), respectively. In this regard, it must be noted that, unlike in the experiment by Slovic *et al.* (1990), the differences that we observed among the three groups cannot be explained by the compatibility between the outcomes of the lotteries and the units in which the certainty equivalent (CE) values for the gambles were stated, since in all groups both the outcomes and the CE values were expressed in the same scale –i.e. in euro (Group 1), number of years in good health (Group 2) and number of days without back pain (Group 3).

The greater rates of SPR in Group 2 than in the other two groups might be explained because in the former group the failure outcome was death. Given that individuals' risk aversion is possibly enhanced when they deal with lotteries that entail a certain risk of death (as those used in Group 2), it is understandable that, in the choices, respondents belonging to Group 2 exhibited a stronger preference for the P-bets than those in the other two groups. This suggests that individuals in Group 2 were more risk averse than the rest of respondents when choosing. Notwithstanding, no clear pattern was observed between Group 2 and the other two groups in relation to the attitude towards risk that respondents displayed in the CE questions (as assessed by the difference between the reported CE values and the expected values of the gambles).

Although lower, the proportions of SPR obtained in Group 3 were also substantial. Thus, the high rates of SPR obtained in Groups 2 and 3 highlights the need to carefully design those studies in which preferences over health outcomes are elicited, trying to minimise the presence of PR.

As far as Group 1 is concerned, it must be borne in mind that, although this group presented the lowest proportions of SPR in the two sessions, it was also the group with the highest rates of NSPR. Overall, the consistency rate was higher in Group 1 than in Group 2, but lower than in Group 3. It is worth remembering that respondents in Group 1 received monetary incentives in the second session. In this respect, some of our results seem to suggest that the incentives had a moderate effect on the stated preferences of subjects in Group 1. This agrees with prior studies in which the use of monetary incentives has not been found to have a significant effect on PR (Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Camerer and Hogarth, 1999).

Our hypothesis regarding the starting point value presented in the CE questions was not confirmed. Indeed, the results obtained in the second session contradicted our expectations, in that the low starting point format gave rise to the highest proportions of SPR (and, in consequence, to the lowest rates of consistent responses), whereas the high starting point format led to the lowest percentages of SPR (and to the highest consistency rates).

One of the most intriguing findings of this study was that women incurred in SPR more frequently than men. This gender difference was robust to the acquisition of experience, since it persisted in all rounds. The fact that women were more likely than men to incur in SPR is primarily explained because the former group chose the P-bets at a significantly higher rate than the latter –suggesting that women were more risk averse than men when choosing– and, to a lesser extent, because women tended to state lower values for the P-bets and higher values for the $-bets than those reported by men. However, it must be noted that the difference between men and women in regards to their attitudes towards risk was not clearly reflected when comparing their stated CE values for the $-bets with the expected values of those lotteries. The stronger risk aversion that women exhibited in the choice tasks is consistent with previous research (Schubert *et al.*, 1999; Hartog *et al.*, 2002; Eckel and Grossman, 2008) and could be ascribable to various factors. For instance, psychologists have observed that women tend to experience negative outcomes more strongly than men (Fujita *et al.*, 1991). In consequence, compared with men, women are more susceptible to overweight the probability of loss (Grossman and Wendy, 1993). Moreover, it has been found that women are less confident under uncertain situations than men, and that women often perceive risky situations as threats, whilst men generally view them as challenges (Lichtenstein *et al.*, 1982; Lundeberg *et al.*, 1994).

Of course, the gender differences we observed might be attributable to the use of different decision processes. In this regard, it has been demonstrated that, even though men and women have similar average levels of general intelligence, they use their brain differently (Zaidi, 2010). For instance, it has been found that the inferior-parietal lobule –the area of the brain that governs numerical brain functions, among others– is significantly larger in men than in women (Frederikse *et al.*, 1999). This fact has been used to explain why men usually outperform women in quantitative tasks, such as mathematical tests (Wilson, 1992; Falk, 2005). Along these lines, since the CE questions are basically quantitative tasks, it is possible that women's valuations were less accurate than men's and/or that women had more difficulties to state their CE values for the gambles. This presumption is supported by the fact that, in both sessions, women required more time to complete the survey, on average. Furthermore, women's valuations displayed higher dispersion than men's, which may suggest that women were more imprecise or less certain than men, at least in the CE questions. In this regard, imprecision in decision making has been found to be associated with a higher likelihood of PR (Butler and Loomes, 2007).

The comparison of our results with prior experiments that have also tested for gender differences in PR is limited because, apart from being scarce, they cast equivocal conclusions. For example, Drichoutis *et al.* (2010) observed that, in mixed gender sessions, men were more likely to incur in PR than women. However, women committed PR more frequently in single gender sessions than in mixed gender sessions, while men exhibited the opposite pattern. In another PR experiment where the lotteries were defined as income distributions of hypothetical countries (Amiel *et al.*, 2008), it was found that the gender differences depended on the question order. In design 1 –where the $-bet (i.e. country 'Alfaland') was valued before the P-bet (i.e. country 'Betaland')– the rate of SPR was higher in the male group than in the female group, whereas the reverse pattern was observed in design 2 –where the P-bet was valued before the $-bet.

Below we will point out the main limitations of the study, along with some suggestions for future research. First, despite the significant downward trend that SPR experienced over the second session (with repetition and feedback), the final proportion of SPR was still unacceptably high so as to consider it to be satisfactory. This high rate could be explained because the CE questions (and the choices in the first session) were repeated a limited number of times (three times for each lottery) and, in consequence, there might not be a real learning effect. Perhaps with more rounds involving repetition and feedback the decay in SPR would have been much more pronounced. Similarly, the relatively small number of repetitions could also explain the upward trend in NSPR that we observed over the second session. It is also possible that this pattern simply arose as a result of response error and and/or by respondents' imprecision.

Likewise, although there is little to learn about a choice task, choices might be influenced by experience too. For example, through repeated choices subjects can better understand the meaning of probabilities. In this respect, it could be conjectured that the results of the second session might have been significantly different if choices had been repeated in that session. However, this presumption does not seem very plausible, since there was generally little variation in choice behaviour over the first session. Furthermore, no significant changes in choice behaviour have been observed in experiments entailing repetition of choices (Butler and Loomes, 2007; Braga *et al.*, 2009). For instance, in the second experiment reported by Braga *et a*l. (2009), there was considerable switching of individual responses across repeated choice tasks, but this variation was apparently random, because it was cancelled out at the aggregate level. Therefore, it seems that the evolution of PR in that study was unaffected by switching of choices.

Another possible limitation is that in the choice tasks indifference between the P-bet and the $-bet was not allowed, as commonly done in choice experiments (Tversky *et al.*, 1990; Braga *et al.*, 2009). In that way, we tried to prevent a large number of false indifferent choices, although we are aware that the fact of having not included the option 'indifferent' in the choices might force some respondents who were actually indifferent between the two paired bets to select at random one of them.[19]

There are also some limitations in relation to the stimuli through which learning was fostered. For instance, in the second session feedback was provided through the resolution of risk of the gambles. This was done by means of a visual aid that showed respondents how the lottery that they had just valued could yield either gains or losses, depending on the probabilities

[19] We assumed that respondents who stated the same value for the two paired bets were indifferent between the two alternatives. However, in view of other researchers that have permitted indifference in the choices (e.g. Oliver, 2006, 2013a), those who place the same value on the two paired gambles and choose the P-bet commit a 'weak predicted reversal' (as well as those who are indifferent in the choices but attach a higher value to the $-bet than to the P-bet), those who value the two lotteries equally and choose the $-bet incur in a 'weak unpredicted reversal' (as well as those who are indifferent in the choices but state a higher value for the P-bet than for the $-bet), whilst those who ascribe the same value to both bets and are indifferent between them in a choice task are consistent with the economic theory.

of success and failure of that gamble. A limitation of this type of visual aid is that it did not show the impact of the different treatments on quality of life, but only the risk associated with them. There are other kinds of visual aids for decision making that may convey more meaningful information, for example, videos, leaflets or computer/web applications in which the impact of the treatments under consideration on patients' quality of life is shown.

With respect to the participants in the experiment, since we used a convenience sample, being relatively homogeneous in terms of age and educational background (most respondents studied Business, Economics or Marketing), our results should be treated with caution and no generalisations can be made from them. However, it is worth noting that experiments on PR usually involve convenience samples and that the sample size in our experiment was quite larger than in most previous PR studies. Additionally, we cannot rule out the possibility that some respondents provided biased responses, for instance, as a result of tiredness, framing effects, the adoption of heuristics or rules of thumb.

To conclude, our results should encourage further research. For instsance, an open question that well deserves to be answered in future studies is whether there is a certain number of repetitions beyond which choice and valuation behaviours –and, consequently, the presence of PR– become stable. It would be also interesting to examine if the fact of allowing respondents to be indifferent in the choices leads to significantly different results from those obtained when this option is not permitted. Moreover, other studies could test, for example, whether the combination of repetition and feedback also contributes to erode other forms of PR that have been identified within the field of health economics (such as judgment-judgment reversals and choice-choice reversals). Furthermore, the fact that women were more likely to incur in SPR throughout the experiment should also stimulate additional research on gender differences in relation to PR, deepening into biological aspects such as hormonal and neural differences between men and women. Finally, if procedural invariance is systematically violated, then future studies should try to get a more thorough insight into individuals' true preferences. For instance, it might be useful to ask respondents to provide qualitative explanations of their responses. Alternatively, the use of more sophisticated techniques (such as the functional magnetic resonance imaging) could contribute to discover subjects' underlying preferences by examining brain activation when they assess different alternatives.

4.5. APPENDIX 4A

The questionnaires

PRESENTACIÓN

Este estudio pretende profundizar en el conocimiento de las preferencias de los ciudadanos en situaciones donde las decisiones están sujetas a incertidumbre. Para conseguir este objetivo le pediremos que elija entre diferentes alternativas cuyas consecuencias no son seguras, así como que les asigne un valor monetario. En cualquiera de los dos casos pretendemos que la única base de tales decisiones sea exclusivamente sus propias preferencias. En este sentido, es importante subrayar que no hay preferencias correctas o incorrectas, lo importante es que sean sus verdaderas preferencias, aquello que realmente piensa que es mejor para usted.

Muchas gracias por su colaboración.

GRUPO 1 (“DINERO”), SESIÓN 1

1) EQUIVALENTE DE CERTEZA:

1.a. Equivalente de certeza inicial “0”

Tiene ante usted dos opciones que comportan la posibilidad de ganar un determinado premio en metálico:

- Alternativa 1: 95 de cada 100 personas que eligen esta opción ganan 8 euros, mientras 5 de cada 100 personas no ganan nada.
- Alternativa 2: 100 de cada 100 personas que eligen esta opción ganan M euros.

A la vista de estas dos opciones, le pedimos que, por favor, establezca la suma segura de euros M para la cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, la cantidad de euros que haría que ambas alternativas fuesen igualmente preferidas para usted.
Inicialmente M se ha fijado en “0 euros”. Usted puede incrementar esa cantidad inicial presionando la flecha “↑”. Si, tras presionar dicho botón, usted alcanzase una suma de euros excesiva de acuerdo a sus preferencias, puede reducir dicho valor presionando el botón “↓”. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón “Siguiente”.
Una vez haya fijado definitivamente la cantidad de euros M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione “Siguiente”.

1.b. Equivalente de certeza inicial “8”

[Todo igual salvo que el quinto párrafo de la tarea dice:
“Inicialmente M se ha fijado en “8 euros”. Usted puede reducir esa cantidad inicial presionando la flecha “↓”. Si, tras presionar dicho botón, usted alcanzase una suma de euros demasiado baja de acuerdo a sus preferencias, puede incrementar dicho valor presionando el botón “↑”. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón “Siguiente”].

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el quinto párrafo de la tarea dice:
“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presione el botón “Siguiente”].

[A CONTINUACIÓN HABRÍA TRES EQUIVALENTES DE CERTEZA MÁS; TANTOS COMO LOTERÍAS RESTAN PARA COMPLETAR LAS CUATRO QUE HAY QUE VALORAR. EL ORDEN DE APARICIÓN SE ALEATORIZA]

…

2) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS

Tiene ante usted dos opciones que comportan la posibilidad de ganar un determinado premio en metálico:

- Alternativa A: 95 de cada 100 personas que eligen esta opción ganan 8 euros, mientras 5 de cada 100 personas no ganan nada.
- Alternativa B: 25 de cada 100 personas que eligen esta opción ganan 27 euros, mientras75 de cada 100 personas no ganan nada.

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere la Alternativa A no tiene más que marcar la respuesta “Prefiero la Alternativa A”. Si, en cambio, usted prefiere la Alternativa B entonces marque la respuesta “Prefiero la Alternativa B”.

Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

3) SEGUNDA Y ÚLTIMA ELECCIÓN ENTRE LOTERÍAS

[Todo es igual que en la pantalla 2) salvo que ahora las alternativas son la C y la D]
[RECORDAR QUE EL ORDEN DE APARICIÓN DE LAS ELECCIONES SE ALEATORIZA].

[RONDA 6: EXPLICACIÓN DEL SISTEMA DE INCENTIVOS]

Para terminar, vamos a explicarle cómo funciona el mecanismo que en la próxima sesión servirá para determinar cuánto dinero puede ganar aparte de la gratificación fija de 10 € que usted percibirá por participar en este estudio.

EJEMPLO GUIADO

PANTALLA 1

Como ha tenido oportunidad de comprobar, en varias ocasiones a lo largo de esta sesión le hemos pedido que determinase la suma segura de dinero M que usted considera equivalente a una determinada alternativa que le ofrece una cierta probabilidad de ganar una cantidad de dinero y una cierta probabilidad de no ganar nada. En otras ocasiones, en cambio, le hemos pedido que eligiese entre dos alternativas.

Durante la próxima sesión volveremos a formularle preguntas semejantes a estas, de modo que al término de la misma seleccionaremos al azar una de ellas, determinando su premio a partir de la respuesta que usted dio a dicha pregunta.

Para que usted pueda comprender claramente la mecánica del procedimiento que permitirá determinar el premio máximo que usted podrá obtener por participar en este experimento, vamos a ver varios ejemplos. En los primeros ejemplos le guiaremos en sus respuestas; posteriormente los hará usted solo. Finalmente le haremos un par de preguntas para comprobar que ha entendido adecuadamente cómo funciona este sistema.

PANTALLA 2

Una vez haya completado todas las preguntas planteadas a lo largo de la sesión, el ordenador elegirá al azar una de ellas, para lo cual aparecerá un bombo, como el de la lotería, en pantalla.

Pulse el botón “Siguiente”

PANTALLA 3

[Aparece el bombo dando vueltas y cae una bola con el número “5”]

En este ejemplo, el ordenador ha seleccionado la pregunta número 5 que, suponga, consistió en que usted eligiese la alternativa más preferida de las dos opciones siguientes:

- Alternativa 1: 90 de cada 100 personas que eligen esta opción ganan 10 euros, mientras 10 de cada 100 personas no ganan nada.
- Alternativa 2: 15 de cada 100 personas que eligen esta opción ganan 30 euros, mientras que 65 de cada 100 personas no ganan nada.

Suponga ahora que usted eligió la alternativa 1.

Por favor, pulse el botón “Siguiente” para continuar

PANTALLA 4

Como usted eligió la alternativa 1, el máximo premio que usted ganaría (aparte de los 10 € fijos) sería el resultado de “jugar” dicha alternativa.

Por favor, pulse el botón “Siguiente” y le explicaremos cómo se jugará la alternativa.

PANTALLA 5: PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

Tiene ante usted 20 puertas cerradas. Tras cada una de esas puertas hay 5 figuras humanas. Las 100 figuras se han distribuido entre las 20 puertas de acuerdo a las probabilidades de la Alternativa 1, esto es: 90 de cada 100 personas ganan 10 euros, mientras que 10 de cada 100 no ganan nada. Hay, por tanto, 18 puertas (= 90/5) tras las que hay personas que ganan 10 euros (figuras que colorearemos de verde) y 2 puertas (= 10/5) tras las que hay personas que no ganan nada (coloreadas de rojo). Imagine que 1 de esas 100 figuras humanas es usted.

Tras presionar el botón “Resolver” usted presenciará cómo se resuelven las probabilidades que encierra la Alternativa 1. Si la puerta que finalmente se abra tiene en su interior 5 figuras de color verde significará que usted ha ganado 10 euros. Si en cambio tras la puerta que se abra hay 5 figuras de color rojo, significará que usted no ha ganado nada.

Por favor, pulse el botón “Resolver” y sabrá cuánto podría ganar.

PANTALLA 6

[Tras el desplazamiento del recuadro coloreado a lo largo de las diferentes puertas, se para en una que al abrirse muestra cinco figuras humanas coloreadas de verde y aparece en pantalla un mensaje en letras grandes que dice: ¡Ha ganado 10 euros!]

PANTALLA 7

Vamos ahora a explicarle cómo funciona este procedimiento para el caso en que la pregunta seleccionada al azar fuese una de las tareas en las que le pedimos que determinase la suma segura de dinero M que usted considera equivalente a una determinada alternativa probabilística como, por ejemplo, la Alternativa 1.

Una vez llegue al término de la segunda sesión aparecerá, como ya le indicamos en el ejemplo anterior, el bombo que indicará qué tarea de todas las que haya completado servirá para determinar su remuneración total. Suponga ahora que el número de la pregunta mostrada en la bola extraída del bombo corresponde a una tarea que consistió en que usted estableciese la suma segura de euros M que hacía que usted considerase equivalentes (igual de buenas) las dos alternativas siguientes:

- Alternativa 1: 90 de cada 100 personas que eligen esta opción ganan 10 euros, mientras 10 de cada 100 personas no ganan nada.
- Alternativa 2: 100 de cada 100 personas que eligen esta opción ganan M euros.

Imagine que usted fijó dicha suma M en 7 euros. Esto es, usted considera que la suma de 7 euros es equivalente a la Alternativa 1, de modo que no sabría por cuál de las dos decantarse.

Por favor, pulse el botón “Siguiente” para continuar.

PANTALLA 8

A continuación, cuando usted pulse el botón “Generar oferta aleatoria”, el ordenador generará un precio por la Alternativa 1 que usted ha considerado equivalente a ganar con seguridad 7 euros.

Si dicho precio resulta menor que la cuantía M fijada (5 euros en este ejemplo) entonces procederemos a jugar la Alternativa 1, esto es, a resolver sus probabilidades como hicimos antes en el ejemplo anterior, de modo que usted podría ganar 10 euros o bien no ganar nada, de ser seleccionada esta tarea. Si por el contrario el precio generado resulta igual o mayor que la cuantía M (7 euros) entonces usted recibiría dicho precio, de ser finalmente seleccionada esta tarea.

Tras presionar el botón “Generar oferta aleatoria” usted presenciará qué cantidad podría ganar de ser ésta la tarea elegida al final de la próxima sesión. Utilizaremos nuevamente un bombo para mostrarle en pantalla el precio aleatorio que genera el ordenador.

PANTALLA 9

[Aparece el bombo dando vueltas y cae una bola que muestra “8 euros”]

Así pues, en este ejemplo el precio generado al azar es de 8 euros. Como es un precio mayor que el que usted consideró equivalente a la Alternativa 1 (7 euros), usted lo recibiría como premio al término de la segunda sesión, de ser ésta la tarea elegida al azar por el ordenador. Si por el contrario la oferta generada aleatoriamente hubiera sido inferior a 7 euros, entonces usted no habría “vendido” la alternativa A, recibiendo como premio el que resultase de jugar la Alternativa 1.

PANTALLA 10

Es fundamental que entienda que lo que más le conviene siempre es revelar la verdadera cantidad M que usted considera equivalente a la alternativa que corresponda. Como el precio ofrecido por el ordenador es generado por completo al azar, es totalmente independiente de la suma M que usted establezca. Fijar una cantidad M diferente a la que usted considera que realmente es equivalente a la alternativa de que se trate sólo puede perjudicarle. Imagine, por ejemplo, que usted piensa que realmente la alternativa A vale 6 euros, pero que usted dice 7 euros, con la esperanza de influir en la oferta del ordenador y venderla a un precio mayor. Si la oferta generada por el ordenador está comprendida entre su verdadero M (es decir, 6 euros) y el que usted declaró (7 euros), usted jugará la lotería, cuando en realidad lo que usted deseaba era venderla. Si usted hubiese declarado 6 euros (el valor que realmente pensaba que tenía la alternativa), habría conseguido venderla. Imagine ahora que usted realmente piensa que la alternativa vale 7 euros, pero dice 6 euros con la esperanza de que así le será más fácil venderla por un precio mayor. Si la oferta generada por el ordenador resulta mayor que 6 euros e inferior a 7, usted malvendería la alternativa cuando en realidad lo que usted deseaba era no deshacerse de ella por menos de 7 euros. Si usted hubiese declarado su verdadero valor equivalente (7 euros) usted no habría vendido la alternativa y podría haber llegado a ganar 10 euros que es premio máximo que ofrece la Alternativa 1 con un 90% de probabilidad.

En definitiva, lo que más le conviene siempre es ser sincero y revelar el valor M que realmente considera alternativo a la alternativa en cuestión.

El siguiente ejemplo no será guiado, de modo que usted debe responder en cada caso lo que auténticamente piensa. Los procedimientos aleatorios que se aplicarán serán totalmente reales, de modo que no está para nada predeterminado el resultado final.

EJEMPLO NO GUIADO

Tiene ante usted dos opciones que comportan la posibilidad de ganar un determinado premio en metálico:

- Alternativa 1: 15 de cada 100 personas que eligen esta opción ganan 30 euros, mientras que 65 de cada 100 personas no ganan nada.
- Alternativa 2: 100 de cada 100 personas que eligen esta opción ganan M euros.

A la vista de estas dos opciones, le pedimos que, por favor, establezca la suma segura de euros M para la cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, la cantidad de euros que haría que ambas alternativas fuesen igual de buenas para usted.

Inicialmente M se ha fijado en “0 euros”. Usted puede incrementar esa cantidad inicial presionando la flecha “↑”. Si, tras presionar dicho botón, usted alcanzase una suma de euros excesiva de acuerdo a sus preferencias, puede reducir dicho valor presionando el botón “↓”. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón “Siguiente”.

Una vez haya fijado definitivamente la cantidad de euros M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione “Siguiente”.

[A PARTIR DE AQUÍ SE DESARROLLA EL MECANISMO BDM: ES DECIR, TRAS FIJAR M Y PULSAR “SIGUIENTE” SE APLICAN LAS PANTALLAS 8 Y 9, CON LAS LÓGICAS ADAPTACIONES A LA CANTIDAD QUE HAYA FIJADO EL SUJETO]

PREGUNTAS PARA COMPROBAR SI SE ENTIENDE EL SISTEMA

Para terminar, le haremos ahora un par de preguntas para asegurarnos de que ha entendido la mecánica del procedimiento que le hemos expuesto.

Pregunta 1.- Suponga que usted ha indicado que considera que 9 euros es una suma equivalente a la Alternativa 1: 95 de cada 100 personas que eligen esta opción ganan 12 euros, 5 de cada 100 no ganan nada.

¿Cuál de las siguientes ofertas realizadas al azar por el ordenador implicaría la venta de la Alternativa 1?

a) 8 euros b) 9 euros c) 8,5 euros

Pregunta 2.- Suponga que usted ha indicado que considera que 14 euros es una suma equivalente a la Alternativa 2: 15 de cada 100 personas que eligen esta opción ganan 20 euros, 85 de cada 100 no ganan nada.

¿Cuál de las siguientes ofertas realizadas al azar por el ordenador implicaría jugar la Alternativa 2?

a) 14 euros b) 15 euros c) 10 euros

GRUPO 1 (“DINERO”), SESIÓN 2

1) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO:

1.a. Equivalente de certeza inicial “0”

Tiene ante usted dos opciones que comportan la posibilidad de ganar un determinado premio en metálico:

- Alternativa 1: 95 de cada 100 personas que eligen esta opción ganan 8 euros, mientras 5 de cada 100 personas no ganan nada.
- Alternativa 2: 100 de cada 100 personas que eligen esta opción ganan M euros.

A la vista de estas dos opciones, le pedimos que, por favor, establezca la suma segura de euros M para la cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, la cantidad de euros que haría que ambas alternativas fuesen igualmente buenas para usted.

Inicialmente M se ha fijado en “0 euros”. Usted puede incrementar esa cantidad inicial presionando la flecha “↑”. Si, tras presionar dicho botón, usted alcanzase una suma de euros excesiva de acuerdo a sus preferencias, puede reducir dicho valor presionando el botón “↓”. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón “Siguiente”.

Una vez haya fijado definitivamente la cantidad de euros M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione “Siguiente”.

1.b. Equivalente de certeza inicial “8”

[Todo igual salvo que el quinto párrafo de la tarea dice:

“Inicialmente M se ha fijado en “8 euros”. Usted puede reducir esa cantidad inicial presionando la flecha “↓”. Si, tras presionar dicho botón, usted alcanzase una suma de euros demasiado baja de acuerdo a sus preferencias, puede incrementar dicho valor presionando el botón “↑”. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón “Siguiente”].

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el quinto párrafo de la tarea dice:

“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de euros M de la Alternativa 2 no será definitivo hasta que usted presione el botón “Siguiente”].

2) PANTALLA DE TRANSICIÓN

Hay estudios científicos que sugieren que la población encuentra difícil manejar probabilidades relacionadas con la ganancia y pérdida de sumas monetarias, como las que usted acaba de observar. Algunos de estos estudios indican que las personas parecen comprender mejor el significado de esas probabilidades si son capaces de visualizar cómo las mismas pueden conducir a diferentes resultados cuando se resuelve el riesgo que conllevan de forma reiterada. A continuación, aplicaremos este enfoque de resolución del riesgo para ayudarle a ajustar la cantidad de euros M que hace equivalentes a su juicio a las alternativas 1 y 2.

Pulse “Siguiente” para continuar.

3) PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

Tiene ante usted 20 puertas cerradas. Tras cada una de esas puertas hay 5 figuras humanas. Las 100 figuras se han distribuido entre las 20 puertas de acuerdo a las probabilidades de la Alternativa 1, esto es: 95 de cada 100 personas ganan 8 euros, mientras que 5 de cada 100 no ganan nada. Imagine que 1 de esas 100 figuras humanas es usted.

Tras presionar el botón “Siguiente” usted presenciará cómo se resuelven las probabilidades que encierra la Alternativa 1 en 10 ocasiones consecutivas. En unas ocasiones el resultado de dicha resolución sería ganar 8 euros, en otros casos, sin embargo, puede ocurrir que el resultado sea no ganar nada. A lo largo de todo el proceso usted podrá comprobar la frecuencia con que ocurre cada una de las posibles consecuencias de la Alternativa 1: bien ganar 8 euros, bien ganar 0 euros.

Por favor, pulse “Siguiente” para comenzar las repeticiones.

[A lo largo del proceso de resolución ha de permanecer visible en pantalla una tabla en la que se vaya acumulando la frecuencia con que acontece cada uno de los resultados de la Alternativa 1, mostrando además el equivalente de certeza que fijó el sujeto así como el premio que finalmente obtendría de elegirse finalmente la tarea a la conclusión de la sesión. Es decir debería ser algo así:

Alternativa 1		Suma segura M
8 euros	0 euros	Usted dijo:
X veces	Y veces	M euros

4) EQUIVALENTE DE CERTEZA DESPUÉS DE LA PRIMERA TANDA DE RESOLUCIONES DEL RIESGO:

A la vista de la frecuencia con que han ocurrido cada uno de los dos posibles resultados de la Alternativa 1, ganar 8 euros y ganar 0 euros [QUE LA TABLA EN LA QUE SE HAN IDO MOSTRANDO LOS RESULTADOS, ASÍ COMO EL EQUIVALENTE DE CERTEZA INICIAL Y EL PREVIO A GANAR PERMANEZCAN EN PANTALLA], vuelva por favor a establecer establezca la suma segura de euros M para la cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, la cantidad de euros que haría que ambas alternativas fuesen igualmente buenas para usted.

El procedimiento para fijar M es el mismo que antes. Recuerde que usted puede incrementar la cantidad inicial de 0 euros presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase una suma de euros excesiva de acuerdo a sus preferencias, puede reducir dicho valor presionando el botón "↓". El número de euros M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente". Por supuesto, usted puede volver a establecer la misma suma M que fijó anteriormente.

Una vez haya fijado definitivamente la cantidad de euros M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

5) NUEVAMENTE PANTALLA DE TRANSICIÓN

6) NUEVAMENTE PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

[Todo es igual aquí, salvo que el valor del equivalente de certeza que debe mostrar la tabla es el que salga de la pantalla 4) y no de la 1)]

7) EQUIVALENTE DE CERTEZA TRAS LA SEGUNDA Y ÚLTIMA TANDA DE RESOLUCIONES DEL RIESGO

[Todo es igual aquí, salvo que la tabla mostrada sería la derivada de la pantalla 6)]

8) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO PERO DE UNA NUEVA LOTERÍA

… [ASÍ CONTINÚAN LOS EQUIVALENTES DE CERTEZA HASTA HABER CUBIERTO LAS CUATRO LOTERÍAS]

…

9) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS

Tiene ante usted dos opciones que comportan la posibilidad de ganar un determinado premio en metálico:

- Alternativa A: 95 de cada 100 personas que eligen esta opción ganan 8 euros, mientras 5 de cada 100 personas no ganan nada.
- Alternativa B: 25 de cada 100 personas que eligen esta opción ganan 27 euros, mientras75 de cada 100 personas no ganan nada.

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere la Alternativa A no tiene más que marcar la respuesta “Prefiero la Alternativa A”. Si, en cambio, usted prefiere la Alternativa B entonces marque la respuesta “Prefiero la Alternativa B”.

Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

10) SEGUNDA ELECCIÓN ENTRE LOTERÍAS

[Todo es igual que en la pantalla 8) salvo que ahora las alternativas son la C y la D]

GRUPO 2 (“AÑOS DE VIDA”), SESIÓN 1

1) EQUIVALENTE DE CERTEZA:

1.a. Equivalente de certeza inicial “0”

Imagine que padece una grave enfermedad, de modo que si no recibe urgentemente un tratamiento usted morirá en unas pocas semanas.

Hay dos posibles tratamientos que pueden curarle, aunque sus efectos sobre su esperanza de vida son distintos:

- Tratamiento 1: 95 de cada 100 personas que eligen esta opción viven 8 años en buena salud, mientras 5 de cada 100 personas fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como el resultado “0 años de vida”).
- Tratamiento 2: 100 de cada 100 personas que eligen esta opción disfrutan de M años de vida en buena salud.

A la vista de estas dos opciones, le pedimos que, por favor, establezca el número de años de vida M para el cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de años de vida que haría que ambas alternativas fuesen igualmente preferidas para usted.

Inicialmente M se ha fijado en “0 años de vida”. Usted puede incrementar esa cantidad inicial presionando la flecha “↑”. Si, tras presionar dicho botón, usted alcanzase un número de años de vida excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón “↓”. El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presiones el botón “Siguiente”.

Una vez haya fijado definitivamente el número de años de vida M de la Tratamiento 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione “Siguiente”.

1.b. Equivalente de certeza inicial “8”

[Todo igual salvo que el sexto párrafo de la tarea dice:

“Inicialmente M se ha fijado en “8 años de vida”. Usted puede reducir esa cantidad inicial presionando la flecha “↓”. Si, tras presionar dicho botón, usted alcanzase un número de años de vida demasiado bajo de acuerdo a sus preferencias, puede incrementar dicho número presionando el botón “↑”. El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presiones el botón “Siguiente”].

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el sexto párrafo de la tarea dice:
“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presione el botón “Siguiente”].

[A CONTINUACIÓN HABRÍA TRES EQUIVALENTES DE CERTEZA MÁS; TANTOS COMO LOTERÍAS RESTAN PARA COMPLETAR LAS CUATRO QUE HAY QUE VALORAR. EL ORDEN DE APARICIÓN SE ALEATORIZA]
…

2) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS

Imagine que padece una grave enfermedad, de modo que si no recibe urgentemente un tratamiento usted morirá en unas pocas semanas.
Hay dos posibles tratamientos que pueden curarle, aunque sus efectos sobre su esperanza de vida son distintos:

- Tratamiento 1: 25 de cada 100 personas que eligen esta opción viven 27 años en buena salud, mientras que 75 de cada 100 personas mueren en pocas semanas (lo cual se representa en el gráfico de abajo como el resultado “0 años de vida”).
- Tratamiento 2: 95 de cada 100 personas que eligen esta opción viven 8 años en buena salud, mientras 5 de cada 100 personas fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como el resultado “0 años de vida”).

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere el Tratamiento 1 no tiene más que marcar la respuesta “Prefiero el Tratamiento 1”. Si, en cambio, usted prefiere el Tratamiento 2 entonces marque la respuesta “Prefiero el Tratamiento 2”.
Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

3) SEGUNDA Y ÚLTIMA ELECCIÓN ENTRE LOTERÍAS
[Todo es igual que en la pantalla 2) salvo que ahora las alternativas son la C y la D]
[RECORDAR QUE EL ORDEN DE APARICIÓN DE LAS ELECCIONES SE ALEATORIZA].

GRUPO 2 ("AÑOS DE VIDA"), SESIÓN 2

1) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO:

1.a. Equivalente de certeza inicial "0"

Imagine que padece una grave enfermedad, de modo que si no recibe urgentemente un tratamiento usted morirá en unas pocas semanas.

Hay dos posibles tratamientos que pueden curarle, aunque sus efectos sobre su esperanza de vida son distintos:

- Tratamiento 1: 95 de cada 100 personas que eligen esta opción viven 8 años en buena salud, mientras 5 de cada 100 personas fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como el resultado "0 años de vida").
- Tratamiento 2: 100 de cada 100 personas que eligen esta opción disfrutan de M años de vida en buena salud.

A la vista de estas dos opciones, le pedimos que, por favor, establezca el número de años de vida M para el cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de años de vida que haría que ambas alternativas fuesen igualmente preferidas para usted.

Inicialmente M se ha fijado en "0 años de vida". Usted puede incrementar esa cantidad inicial presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de años de vida excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presiones el botón "Siguiente".

Una vez haya fijado definitivamente el número de años de vida M del Tratamiento 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

1.b. Equivalente de certeza inicial "8"

[Todo igual salvo que el sexto párrafo de la tarea dice:

"Inicialmente M se ha fijado en "8 años de vida". Usted puede reducir esa cantidad inicial presionando la flecha "↓". Si, tras presionar dicho botón, usted alcanzase un número de años de vida demasiado bajo de acuerdo a sus preferencias, puede incrementar dicho número presionando el botón "↑". El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presiones el botón "Siguiente"].

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el sexto párrafo de la tarea dice:
“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de años de vida M del Tratamiento 2 no será definitivo hasta que usted presione el botón “Siguiente”].

2) PANTALLA DE TRANSICIÓN
Hay estudios científicos que sugieren que la población encuentra difícil manejar probabilidades relacionadas con la vida y la muerte, como las que usted acaba de observar. Algunos de estos estudios indican que las personas parecen comprender mejor el significado de esas probabilidades si son capaces de visualizar cómo las mismas pueden conducir a diferentes resultados cuando se resuelve el riesgo que conllevan de forma reiterada. A continuación, aplicaremos este enfoque de resolución del riesgo para ayudarle a ajustar el número de años de vida M que hace equivalentes a su juicio los tratamientos 1 y 2.
Pulse “Siguiente” para continuar.

3) PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO
Tiene ante usted 20 puertas cerradas. Tras cada una de esas puertas hay 5 figuras humanas. Las 100 figuras se han distribuido entre las 20 puertas de acuerdo a las probabilidades del Tratamiento 1, esto es: 95 de cada 100 personas que eligen esta opción viven 8 años en buena salud, mientras 5 de cada 100 personas fallecen en unas pocas semanas. Imagine que 1 de esas 100 figuras humanas es usted.
Tras presionar el botón “Siguiente” usted presenciará cómo se resuelven las probabilidades que encierra el Tratamiento 1 en 10 ocasiones consecutivas. En unas ocasiones el resultado de dicha resolución sería ganar 8 años de vida en buena salud, en otros casos, sin embargo, puede ocurrir que el resultado sea fallecer en unas pocas semanas. A lo largo de todo el proceso usted podrá comprobar la frecuencia con que ocurre cada una de las posibles consecuencias del Tratamiento 1: bien ganar 8 años de vida, bien morir en pocas semanas.
Piense en cada uno de los resultados que se vayan produciendo como las consecuencias que el Tratamiento 1 ha tenido en otros pacientes similares a usted.
Por favor, pulse “Siguiente” para comenzar las repeticiones.
[A lo largo del proceso de resolución ha de permanecer visible en pantalla una tabla en la que se vaya acumulando la frecuencia con que acontece cada uno de los resultados de la Alternativa 1,

mostrando además el equivalente de certeza que fijó el sujeto así como el premio que finalmente obtendría de elegirse finalmente la tarea a la conclusión de la sesión. Es decir debería ser algo así:

Tratamiento 1		Cantidad segura M
8 años de vida	0 años de vida	Usted dijo:
X veces	Y veces	M años de vida

4) EQUIVALENTE DE CERTEZA DESPUÉS DE LA PRIMERA TANDA DE RESOLUCIONES DEL RIESGO:

A la vista de la frecuencia con que han ocurrido cada uno de los dos posibles resultados del Tratamiento 1, ganar 8 años de vida o fallecer en unas pocas semanas [QUE LA TABLA EN LA QUE SE HAN IDO MOSTRANDO LOS RESULTADOS, ASÍ COMO EL EQUIVALENTE DE CERTEZA INICIAL], vuelva por favor a establecer establezca el número seguro de años de vida M para la cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de años de vida que haría que ambas alternativas fuesen igualmente buenas para usted.

El procedimiento para fijar M es el mismo que antes. Recuerde que usted puede incrementar la cantidad inicial de 0 años de vida presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de años de vida excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de años de vida M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente". Por supuesto, usted puede volver a establecer el mismo número de años de vida M que fijó anteriormente.

Una vez haya fijado definitivamente la cantidad de años de vida M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

5) NUEVAMENTE PANTALLA DE TRANSICIÓN

6) NUEVAMENTE PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

[Todo es igual aquí, salvo que el valor del equivalente de certeza que debe mostrar la tabla es el que salga de la pantalla 4) y no de la 1)] .

7) EQUIVALENTE DE CERTEZA TRAS LA SEGUNDA Y ÚLTIMA TANDA DE RESOLUCIONES DEL RIESGO

[Todo es igual aquí, salvo que la tabla mostrada sería la derivada de la pantalla 6)].

8) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO PERO DE UNA NUEVA LOTERÍA

… [ASÍ CONTINÚAN LOS EQUIVALENTES DE CERTEZA HASTA HABER CUBIERTO LAS CUATRO LOTERÍAS]

…

9) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS

Imagine que padece una grave enfermedad, de modo que si no recibe urgentemente un tratamiento usted morirá en unas pocas semanas.

Hay dos posibles tratamientos que pueden curarle, aunque sus efectos sobre su esperanza de vida son distintos:

- Alternativa 1: 25 de cada 100 personas que eligen esta opción viven 27 años en buena salud, mientras que 75 de cada 100 personas mueren en pocas semanas (lo cual se representa en el gráfico de abajo como el resultado “0 años de vida”).
- Alternativa 2: 95 de cada 100 personas que eligen esta opción viven 8 años en buena salud, mientras 5 de cada 100 personas fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como el resultado “0 años de vida”).

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere la Alternativa A no tiene más que marcar la respuesta “Prefiero la Alternativa A”. Si, en cambio, usted prefiere la Alternativa B entonces marque la respuesta “Prefiero la Alternativa B”.

Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

10) SEGUNDA ELECCIÓN ENTRE LOTERÍAS

[Todo es igual que en la pantalla 9) salvo que ahora las alternativas son la C y la D].

GRUPO 3 ("DÍAS SIN DOLOR DE ESPALDA"), SESIÓN 1

1) EQUIVALENTE DE CERTEZA:

1.a. Equivalente de certeza inicial "0"

Imagine que usted padece de dolor de espalda crónico. Esto significa que sufre dolores y molestias de moderada intensidad la mayor parte del tiempo.

Hay dos medicinas que pueden eliminar el dolor de espalda durante algunos días al mes, en concreto:

- Medicina 1: 95 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 8 días al mes. Sin embargo, el medicamento no tiene ningún efecto en 5 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado "0 días sin dolor").
- Medicina 2: 100 de cada 100 personas que eligen esta opción viven sin dolor ni molestias de espalda durante un cierto número de días al mes M.

A la vista de estas dos opciones, le pedimos que, por favor, establezca el número de días sin dolor de espalda M para el cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de días sin dolor que haría que ambas alternativas fuesen igualmente preferidas para usted.

Inicialmente M se ha fijado en "0 días sin dolor". Usted puede incrementar esa cantidad inicial presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de días sin dolor excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente".

Una vez haya fijado definitivamente el número de días sin dolor M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

1.b. Equivalente de certeza inicial "8"

[Todo igual salvo que el sexto párrafo de la tarea dice:

Inicialmente M se ha fijado en "8 días sin dolor". Usted puede incrementar esa cantidad inicial presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de días sin dolor excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente".

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el sexto párrafo de la tarea dice:
“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presione el botón “Siguiente”].

[A CONTINUACIÓN HABRÍA TRES EQUIVALENTES DE CERTEZA MÁS; TANTOS COMO LOTERÍAS RESTAN PARA COMPLETAR LAS CUATRO QUE HAY QUE VALORAR. EL ORDEN DE APARICIÓN SE ALEATORIZA]

2) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS
Imagine que usted padece de dolor de espalda crónico. Esto significa que sufre dolores y molestias de moderada intensidad la mayor parte del tiempo.
Hay dos medicinas que pueden eliminar el dolor de espalda durante algunos días al mes, en concreto:

- Medicina 1: 25 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 27 días al mes, mientras que el medicamento no tiene ningún efecto en 75 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado “0 días sin dolor”).
- Medicina 2: 95 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 8 días al mes. Sin embargo, el medicamento no tiene ningún efecto en 5 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado “0 días sin dolor”).

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere la Medicina 1 no tiene más que marcar la respuesta “Prefiero la Medicina 1”. Si, en cambio, usted prefiere la Medicina 2 entonces marque la respuesta “Prefiero la Medicina 2”.
Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

3) SEGUNDA Y ÚLTIMA ELECCIÓN ENTRE LOTERÍAS
[Todo es igual que en la pantalla 2) salvo que ahora las alternativas son la C y la D]
[RECORDAR QUE EL ORDEN DE APARICIÓN DE LAS ELECCIONES SE ALEATORIZA].

GRUPO 3 ("DÍAS SIN DOLOR DE ESPALDA"), SESIÓN 2

1) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO:

1.a. Equivalente de certeza inicial "0"

Imagine que usted padece de dolor de espalda crónico. Esto significa que sufre dolores y molestias de moderada intensidad la mayor parte del tiempo.

Hay dos medicinas que pueden eliminar el dolor de espalda durante algunos días al mes, en concreto:

- Medicina 1: 95 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 8 días al mes. Sin embargo, el medicamento no tiene ningún efecto en 5 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado "0 días sin dolor").
- Medicina 2: 100 de cada 100 personas que eligen esta opción viven sin dolor ni molestias de espalda durante un cierto número de días al mes M.

A la vista de estas dos opciones, le pedimos que, por favor, establezca el número de días sin dolor de espalda M para el cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de días sin dolor que haría que ambas alternativas fuesen igualmente preferidas para usted.

Inicialmente M se ha fijado en "0 días sin dolor". Usted puede incrementar esa cantidad inicial presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de días sin dolor excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente".

Una vez haya fijado definitivamente el número de días sin dolor M de la Alternativa 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

1.b. Equivalente de certeza inicial "8"

[Todo igual salvo que el sexto párrafo de la tarea dice:

Inicialmente M se ha fijado en "8 días sin dolor". Usted puede incrementar esa cantidad inicial presionando la flecha "↑". Si, tras presionar dicho botón, usted alcanzase un número de días sin dolor excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓". El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente".

1.c. Equivalente de certeza “abierto”

[Todo igual salvo que el sexto párrafo de la tarea dice:
“Escriba directamente dicha cantidad M utilizando las teclas numéricas del teclado. El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presione el botón “Siguiente”].

2) PANTALLA DE TRANSICIÓN

Hay estudios científicos que sugieren que la población encuentra difícil manejar probabilidades relacionadas con el dolor y las molestias de salud, como las que usted acaba de observar. Algunos de estos estudios indican que las personas parecen comprender mejor el significado de esas probabilidades si son capaces de visualizar cómo las mismas pueden conducir a diferentes resultados cuando se resuelve el riesgo que conllevan de forma reiterada. A continuación, aplicaremos este enfoque de resolución del riesgo para ayudarle a ajustar el número de días sin dolor M que hace equivalentes a su juicio a las medicinas 1 y 2.
Pulse “Siguiente” para continuar.

3) PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

Tiene ante usted 20 puertas cerradas. Tras cada una de esas puertas hay 5 figuras humanas. Las 100 figuras se han distribuido entre las 20 puertas de acuerdo a las probabilidades de la Medicina 1, esto es: 95 de cada 100 personas que toman este medicamento viven sin dolor de espalda 8 días al mes, mientras 5 de cada 100 personas que lo toman no logran alivio ningún día del mes. Imagine que 1 de esas 100 figuras humanas es usted.
Tras presionar el botón “Siguiente” usted presenciará cómo se resuelven las probabilidades que encierra la Medicina 1 en 10 ocasiones consecutivas. En unas ocasiones el resultado de dicha resolución sería disfrutar de 8 días al mes sin dolor ni molestias de espalda, en otros casos, sin embargo, puede ocurrir que el resultado sea no obtener alivio a dicho dolor ni un día siquiera del mes. A lo largo de todo el proceso usted podrá comprobar la frecuencia con que ocurre cada una de las posibles consecuencias de la Medicina 1: bien disfrutar de 8 días al mes sin dolor, bien no obtener alivio a su dolor ningún día del mes.
Piense en cada uno de los resultados que se vayan produciendo como las consecuencias que la Medicina 1 ha tenido en otros pacientes similares a usted.
Por favor, pulse “Siguiente” para comenzar las repeticiones.

[A lo largo del proceso de resolución ha de permanecer visible en pantalla una tabla en la que se vaya acumulando la frecuencia con que acontece cada uno de los resultados de la Medicina 1, mostrando además el equivalente de certeza que fijó el sujeto. Es decir debería ser algo así:

Medicina 1		Cantidad segura M
8 años de vida	0 años de vida	Usted dijo:
X veces	Y veces	M años de vida

4) EQUIVALENTE DE CERTEZA DESPUÉS DE LA PRIMERA TANDA DE RESOLUCIONES DEL RIESGO:

A la vista de la frecuencia con que han ocurrido cada uno de los dos posibles resultados de la Medicina 1, disfrutar de 8 días al mes sin dolor ni molestias de espalda o no lograr alivio ningún día del mes [QUE LA TABLA EN LA QUE SE HAN IDO MOSTRANDO LOS RESULTADOS, ASÍ COMO EL EQUIVALENTE DE CERTEZA INICIAL], vuelva por favor a establecer establezca el número seguro de días sin dolor de espalda M para el cual usted no sabría por cuál de las dos opciones decantarse. En otras palabras, el número de días sin dolor que haría que ambas alternativas fuesen igualmente buenas para usted.

El procedimiento para fijar M es el mismo que antes. Recuerde que usted puede incrementar la cantidad inicial de 0 días sin dolor presionando la flecha "↑".Si, tras presionar dicho botón, usted alcanzase un número de días sin dolor excesivo de acuerdo a sus preferencias, puede reducir dicho número presionando el botón "↓".El número de días sin dolor M de la Alternativa 2 no será definitivo hasta que usted presiones el botón "Siguiente". Por supuesto, usted puede volver a establecer el mismo número de días sin dolor M que fijó anteriormente.

Una vez haya fijado definitivamente el número de días sin dolor M de la Medicina 2 que hace que usted se encuentre indiferente entre las dos opciones, por favor presione "Siguiente".

5) NUEVAMENTE PANTALLA DE TRANSICIÓN

6) NUEVAMENTE PANTALLA EXPLICATIVA DE LA RESOLUCIÓN DEL RIESGO

[Todo es igual aquí, salvo que el valor del equivalente de certeza que debe mostrar la tabla es el que salga de la pantalla 4) y no de la 1)].

7) EQUIVALENTE DE CERTEZA TRAS LA SEGUNDA Y ÚLTIMA TANDA DE RESOLUCIONES DEL RIESGO

[Todo es igual aquí, salvo que la tabla mostrada sería la derivada de la pantalla 6)].

8) EQUIVALENTE DE CERTEZA ANTES DE LA RESOLUCIÓN DEL RIESGO PERO DE UNA NUEVA LOTERÍA

… [ASÍ CONTINÚAN LOS EQUIVALENTES DE CERTEZA HASTA HABER CUBIERTO LAS CUATRO LOTERÍAS]

…

9) PRIMERA ELECCIÓN ENTRE DOS LOTERÍAS

Imagine que usted padece de dolor de espalda crónico. Esto significa que sufre dolores y molestias de moderada intensidad la mayor parte del tiempo.

Hay dos medicinas que pueden eliminar el dolor de espalda durante algunos días al mes, en concreto:

- Medicina 1: 25 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 27 días al mes, mientras que el medicamento no tiene ningún efecto en 75 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado “0 días sin dolor”).
- Medicina 2: 95 de cada 100 personas que toman este medicamento viven sin dolor ni molestias de espalda durante 8 días al mes. Sin embargo, el medicamento no tiene ningún efecto en 5 de cada 100 personas que lo toman (lo cual se representa en el gráfico de abajo como el resultado “0 días sin dolor”).

A la vista de estas dos opciones, le pedimos que, por favor, escoja entre ambas la alternativa que más prefiera. Para ello, si usted prefiere la Medicina 1 no tiene más que marcar la respuesta “Prefiero la Medicina 1”. Si, en cambio, usted prefiere la Medicina 2 entonces marque la respuesta “Prefiero la Medicina 2”.

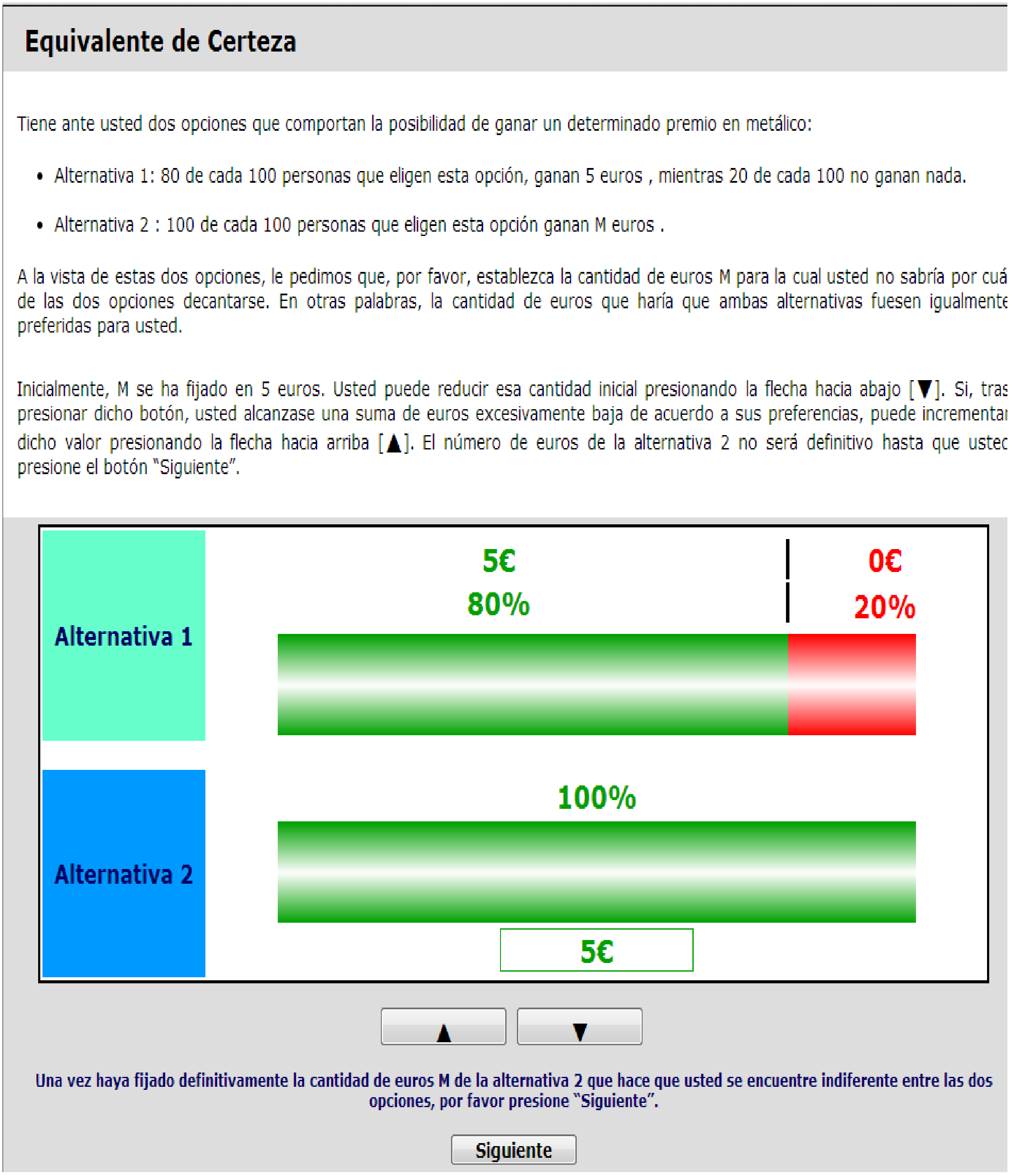
Una vez haya elegido definitivamente la alternativa que más prefiera pulse el botón “Siguiente”.

10) SEGUNDA ELECCIÓN ENTRE LOTERÍAS

[Todo es igual que en la pantalla 9) salvo que ahora las alternativas son la C y la D].

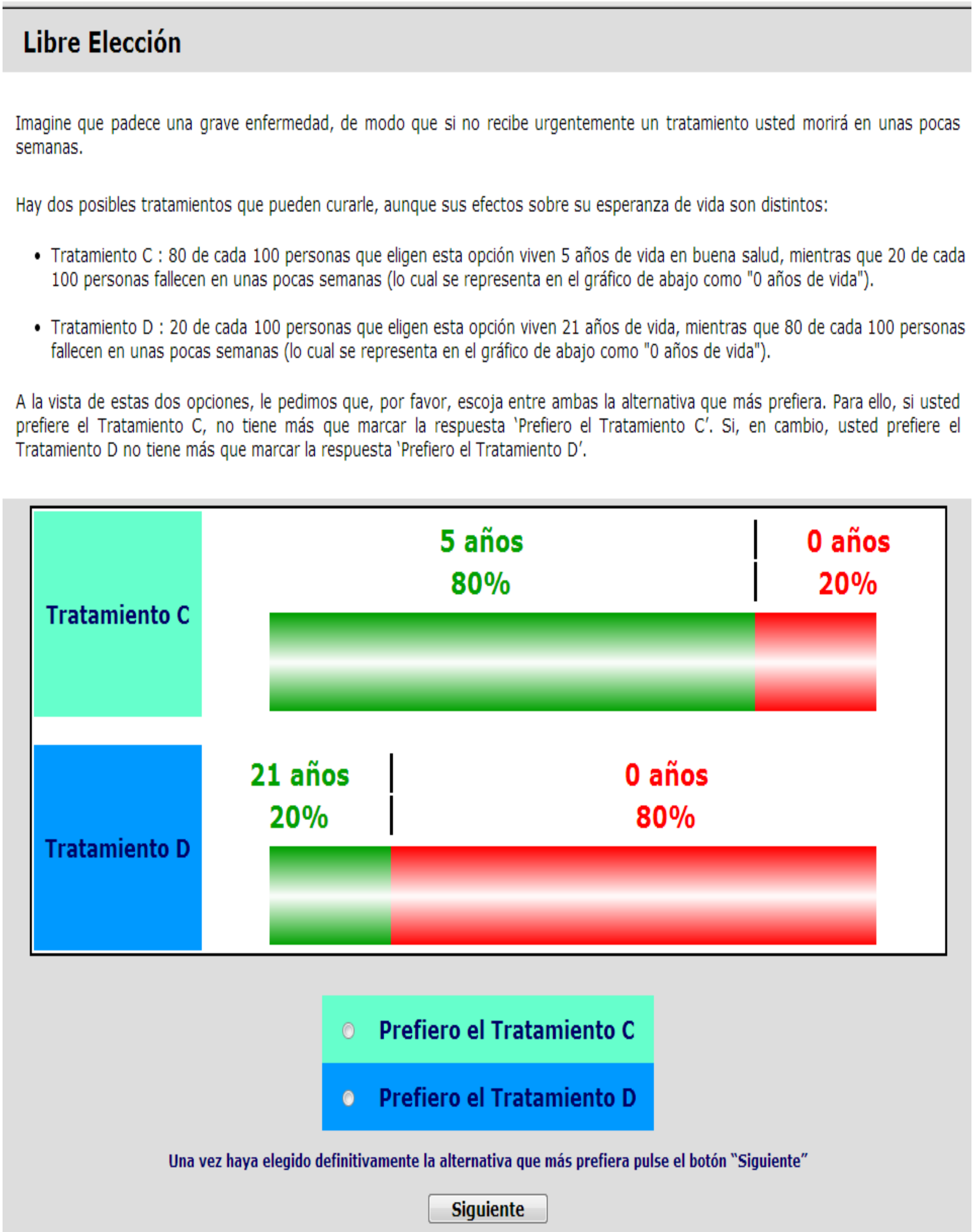
4.6. APPENDIX 4.B

Figure 4B.1. Screenshot of a CE question



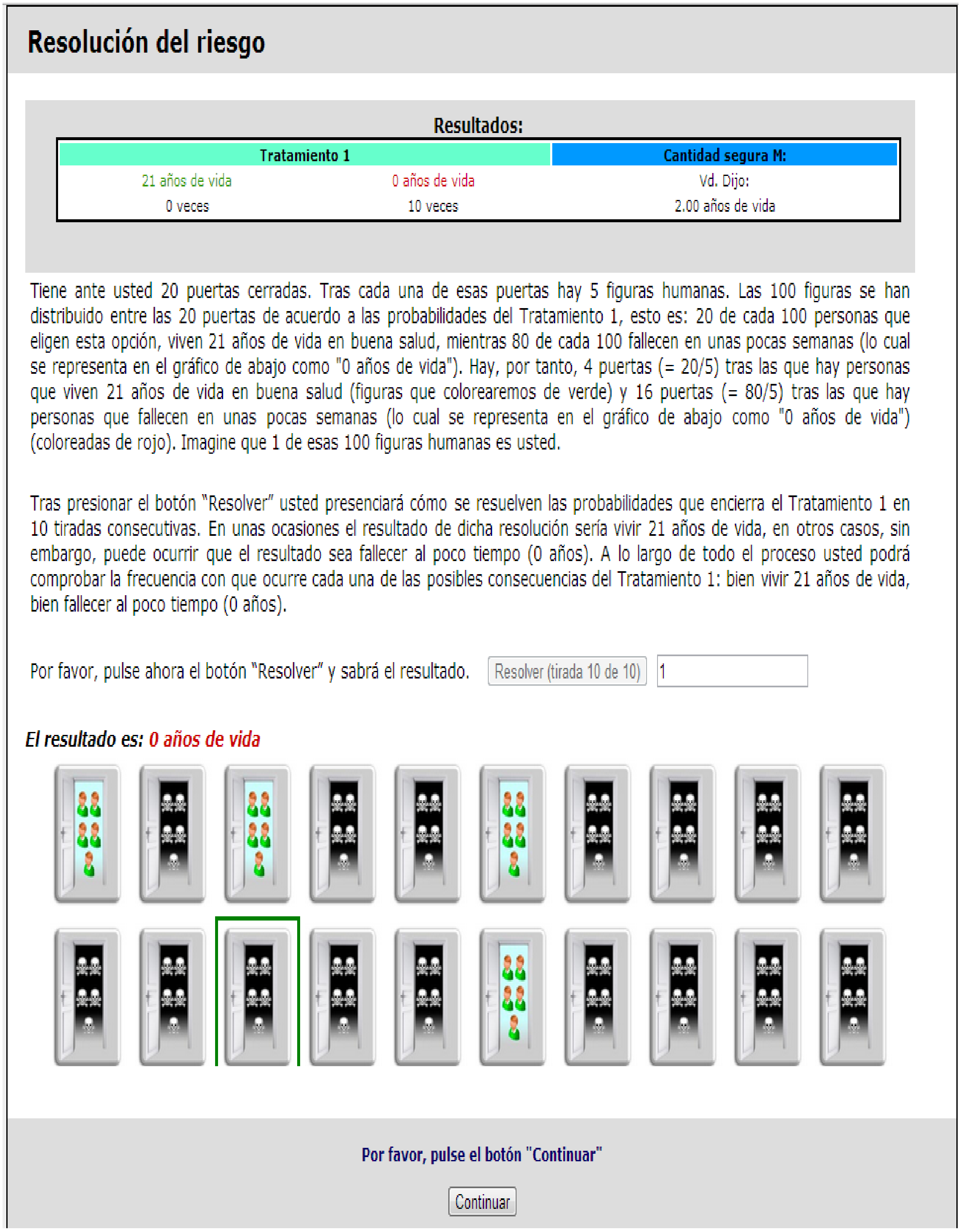
In this example, the bet represented as 'Alternative 1' is lottery C (5, 0.8) and 'Alternative 2' is 'M' (the certain outcome).

Figure 4B.2. Screenshot of a choice question



This example displays a choice between gambles C (5, 0.8) and D (21, 0.2).

Figure 4B.3. Screenshot of the resolution of risk



Resolución del riesgo

Resultados:

Tratamiento 1		Cantidad segura M:
21 años de vida	0 años de vida	Vd. Dijo:
0 veces	10 veces	2.00 años de vida

Tiene ante usted 20 puertas cerradas. Tras cada una de esas puertas hay 5 figuras humanas. Las 100 figuras se han distribuido entre las 20 puertas de acuerdo a las probabilidades del Tratamiento 1, esto es: 20 de cada 100 personas que eligen esta opción, viven 21 años de vida en buena salud, mientras 80 de cada 100 fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como "0 años de vida"). Hay, por tanto, 4 puertas (= 20/5) tras las que hay personas que viven 21 años de vida en buena salud (figuras que colorearemos de verde) y 16 puertas (= 80/5) tras las que hay personas que fallecen en unas pocas semanas (lo cual se representa en el gráfico de abajo como "0 años de vida") (coloreadas de rojo). Imagine que 1 de esas 100 figuras humanas es usted.

Tras presionar el botón "Resolver" usted presenciará cómo se resuelven las probabilidades que encierra el Tratamiento 1 en 10 tiradas consecutivas. En unas ocasiones el resultado de dicha resolución sería vivir 21 años de vida, en otros casos, sin embargo, puede ocurrir que el resultado sea fallecer al poco tiempo (0 años). A lo largo de todo el proceso usted podrá comprobar la frecuencia con que ocurre cada una de las posibles consecuencias del Tratamiento 1: bien vivir 21 años de vida, bien fallecer al poco tiempo (0 años).

Por favor, pulse ahora el botón "Resolver" y sabrá el resultado. Resolver (tirada 10 de 10) 1

El resultado es: 0 años de vida

Por favor, pulse el botón "Continuar"

Continuar

This figure shows an example of the resolution of risk corresponding to lottery D (21, 0.2)

CHAPTER 5

GENERAL CONCLUSIONS

Chapters 2-4 of this thesis have presented three independent empirical studies whose overall aim was to examine new methodological approaches regarding the analysis of costs and outcomes for the economic evaluation of health care interventions. To be exact, Chapter 2 provided estimates of the relationship between health-related quality of life (HRQoL) and life satisfaction by explicitly accounting for the simultaneous association between them; Chapter 3 elicited monetary values for informal care based on both informal carers' and non-carers' preferences; and Chapter 4 tested whether the consistency of preferences (over health outcomes and money) improves as individuals learn through the acquisition of experience with the tasks they make. To conclude with this thesis, the present chapter summarises and discusses the key results of the three studies, points out their main limitations and puts forward suggestions for future research.

5.1. THE SIMULTANEOUS RELATIONSHIP BETWEEN LIFE SATISFACTION AND PREFERENCE-BASED HRQoL

The relationship between health and subjective well-being (SWB) has been extensively investigated, but always under a unidirectional standpoint. For instance, health economists generally conceive that SWB depends on health, which is logical given that the former is a broader construct than the latter. However, there is also wide evidence showing that health is influenced by SWB, thus suggesting that health and SWB are reciprocally related. In Chapter 2 we focused on the relationship between health-related quality of life (HRQoL) and life satisfaction, but, against the unidirectional approaches used in the literature, we examined this association taking a simultaneous perspective (i.e. assuming that HRQoL and life satisfaction influence each other). More specifically, Chapter 2 tried to answer the following questions:

- Is bidirectional the relationship between HRQoL and life satisfaction?
- If so, is there any bias when we model the relationship between them from a unidirectional perspective?
- Which effect is stronger: that from HRQoL to life satisfaction or the opposite one?
- Is the relationship between HRQoL and life satisfaction heterogeneous in terms of age and gender?

To provide an answer to these questions, first we examined the relationship between life satisfaction and HRQoL –as assessed by the Satisfaction With Life Scale (SWLS) summary score and the SF-6D utility index, respectively– under a unidirectional approach (that is, by estimating separate equations for each outcome variable). Next, we compared the resulting unidirectional estimates with those obtained from a simultaneous equations model which accounted for the mutual correlation between life satisfaction and the SF-6D index. The data were collected from a sample of 870 respondents, who were representative of the Spanish adult general public in terms of age and sex.

5.1.1. Main results and discussion

Our results revealed that life satisfaction and HRQoL are simultaneously associated and that the fact of disregarding this simultaneous relationship severely underestimates the effect of life satisfaction on HRQoL (by 18.2%) and, to a lesser degree, the effect of HRQoL on life satisfaction (by 6.3%). We also found the effect of HRQoL on life satisfaction to be stronger than the opposite one, although the difference lowered when accounting for simultaneity. To be exact, under the unidirectional approach, the effect of HRQoL on life satisfaction was 37% higher than the effect of life satisfaction on HRQoL, whilst the coefficients of the simultaneous equations system revealed that the former effect was 19.7% higher than the latter.

Furthermore, we observed age and gender differences in the relationship between HRQoL and life satisfaction. It must be noted that there was no evidence of simultaneity among men and middle-aged respondents (36-49 years). In both cases only the effect from HRQoL to life satisfaction was significant. In the rest of groups –women, younger (< 36 years) and older (50 years and over) individuals– life satisfaction and HRQoL were found to be simultaneously associated and, surprisingly, the effect of life satisfaction on HRQoL turned out to be stronger than the reverse one.

The main contribution of the study reported in Chapter 2 is that it has been the first empirical work to model the relationship between life satisfaction and self-perceived health explicitly accounting for the reciprocal correlation between them. This allowed us to provide more accurate coefficients of the effect of life satisfaction on HRQoL and of the reverse direction effect as compared with the coefficients obtained under the unidirectional approach conventionally adopted in the literature on the relationship between SWB and health.

The fact that HRQoL and life satisfaction are simultaneously related has implications for the economic evaluation of health technologies, for public policy and for health care professionals.

Implications for the economic evaluation

Models of life satisfaction generally include some measure of self-perceived health in the list of regressors, given that health (especially self-perceived health) is one of the most important determinants of SWB. Still, in these models there is a potential risk of simultaneity bias, because the simultaneous association between life satisfaction and self-perceived health is not accounted for. Our results suggest that these models could underestimate the positive effect of self-perceived health on life satisfaction. This potential bias should be borne in mind in future developments of the well-being valuation (WBV) method, for instance. Within the field of health economics, this method has been applied to estimate the necessary monetary amount to compensate the well-being loss caused by certain chronic diseases (Ferrer-i-Carbonell and Van Praag, 2002; Powdthavee and Van den Berg, 2011), as well as to value informal care in monetary terms (Van den Berg and Ferrer-i-Carbonell, 2007). In a similar way, this method could be used to estimate the monetary amount required to compensate HRQoL losses or the monetary value of HRQoL gains. If the WBV method were applied for these purposes and the simultaneous relationship between HRQoL and SWB were ignored, the coefficient of the effect of HRQoL on SWB and, therefore, the resulting monetary valuations would be biased.

On the other hand, models of HRQoL rarely control for life satisfaction (or another variable of SWB), thus neglecting that SWB may have an impact on HRQoL. Chapter 2 showed that this effect is significant and has a considerable size –after adjusting for simultaneity, we found that a 1% increase in life satisfaction leads to a 0.30% increase in the SF-6D index. What is more, the observed positive effect of life satisfaction on HRQoL was stronger than the negative effect on HRQoL caused by any chronic disease included in the regression analysis. More important, we found that the effect of life satisfaction on HRQoL is severely underestimated (by nearly 20%) when simultaneity is not taken into account. In view of the significant effect of life satisfaction on HRQoL found in Chapter 2, we recommend that models of HRQoL control for life satisfaction, along with demographic and socioeconomic factors, chronic diseases and personality traits. Of course, in these models the simultaneous association between HRQoL and life satisfaction should be considered. In this regard, we acknowledge that our simultaneous equations approach can be unfeasible (e.g. the exclusion restrictions may be difficult or impossible to identify). An easiest solution would be the development of corrective weights, able to debias the unidirectional estimates of the effect of life satisfaction on HRQoL and of the opposite effect (e.g. similar to the corrective formulas of prospect theory). To that end, it would be necessary to obtain more evidence on the relationship between HRQoL and life satisfaction (or, more generally, between self-perceived health and SWB). These corrective weights could be applied with both forthcoming and existing estimates reported in previous studies.

An open question that was not addressed in Chapter 2 is how to incorporate both HRQoL and life satisfaction effects in an economic evaluation. It would be wrong to compute HRQoL and life satisfaction changes separately and then to aggregate them to obtain an overall score to be included in the effect side of a cost-utility analysis because it would cause double counting, since life satisfaction is partly determined by HRQoL and vice versa. Another alternative would be to derive a new descriptive system with HRQoL and SWB attributes. A first step could be to take the dimensions of existing generic descriptive systems of HRQoL (such as the EQ-5D and the SF-6D) as starting point and then to add new attributes for the three components of SWB (life satisfaction, positive affect and negative affect), trying to avoid overlaps across HRQoL dimensions and SWB components. In that way, it could be possible to provide more meaningful information about the utility associated with the hypothetical health states being assessed.

Implications for public policy

If non-health policies improve citizens' life satisfaction, then those interventions might boost HRQoL gains through the effect of life satisfaction on HRQoL. Therefore, health care interventions should not be seen as the only instruments to make people healthier, but also non-medical strategies (in areas such as education, employment, the environment, social services, etc.) can be useful tools for the improvement of HRQoL. This is consistent with the ideal of 'health in all policies' (Leppo *et al.*, 2013). This implies that all public policies should be designed and evaluated by examining their consequences on health, although SWB should be still regarded as a broader goal. Thus, under this perspective, all public policies should focus on two main outcome measures: SWB and health. In this respect, the simultaneous equations approach used in Chapter 2 could be helpful for the evaluation of public policies since it allows for the possibility of modelling both HRQoL and life satisfaction outcomes at the same time, taking into account the simultaneous association between them while also controlling for other variables. On the other hand, health care interventions can have effects beyond health (e.g. on overall life satisfaction), which can reinforce the effectiveness of non-medical strategies. In summary, the interplay between HRQoL and life satisfaction highlights the need to coordinate all public policies (health and non-health policies), with the aim of seizing synergies across them.

Implications for health care professionals

The simultaneous relationship between HRQoL and life satisfaction implies that, all else equal, the more satisfied a person is, the more likely he/she will be to experience larger HRQoL gains. Conversely, very unsatisfied people might experience fewer HQoL gains than more satisfied people. Therefore, a given health care intervention (especially those in the area of mental health) could be more cost-effective (at least in the long run) and recovery could be faster if it were supplemented with low-cost non-medical actions that are beneficial for life satisfaction, for example: cognitive therapy (padash *et al.*, 2012), humour therapy (Tse *et al.*, 2010), art therapy (Wilkinson and Chilton, 2013) or volunteering (Corporation for National and Community Service, 2007), among others. In any event, it must be taken into account that it seems unlikely that HRQoL and life satisfaction are synchronised (i.e. that a change in HRQoL probably does not lead automatically to a change in life satisfaction and vice versa). Quite on the contrary, it is logical to think that a change in HRQoL may need some time to have an impact on life satisfaction and vice versa.

5.1.2. Limitations and suggestions for future research

Some limitations of the study presented in Chapter 2 have to be borne in mind. One of them is that, although SWB is made up of three components (satisfaction with life, positive affect and negative affect), we considered only the cognitive component of SWB (i.e. life satisfaction), which is the closest concept to the utilitarian notion of the 'good life' (Veenhoven, 1993). There are several reasons why we focused on satisfaction with life. First, compared with positive and negative affects, life satisfaction is a less ambiguous notion, in the sense that it reflects the discrepancy between aspirations and achievements (Campbell *et al.*, 1976). Furthermore, life satisfaction is more stable than positive and negative affects, which are usually fleeting reactions to specific situations (Pavot and Diener, 1993). It is worth remembering that many researchers have also focused on this component of SWB (Diener *et al.*, 1985; Veenhoven, 1993; Di Tella *et al.*, 2001; Frey and Stutzer, 2002; Blanchflower and Oswald, 2004; Helliwell and Putnam, 2004). In any case, it would be of great interest to investigate whether the other two components of SWB are also simultaneously related to HRQoL and, if so, to compare the size of the simultaneity bias depending on the component of SWB being examined.

By the same token, we used two specific measures to assess life satisfaction and HRQoL (the SWLS summary score and the SF-6D utility index, respectively). In consequence, we cannot discard that other measures would have yielded significantly different results from those reported in Chapter 2. For this reason, our results should be compared with those derived from the use of other HRQoL measures (e.g. the EQ-5D or the HUI3 utility scores) and other life satisfaction measures (e.g. a single question about satisfaction with life as a whole).

Another limitation is that we obtained a static view of life satisfaction and HRQoL because we used cross-sectional data. Given that it is likely that the effect of HRQoL on life satisfaction and the opposite effect do not take place automatically, but they take some time to be observed, the use of panel data would allow us to better understand the dynamics of the relationship between HRQoL and life satisfaction, as well as to obtain more accurate and robust estimates. To that end, future studies on the simultaneous association between self-perceived health and SWB might use existing large longitudinal surveys where both health-related and SWB-related questions are included. Some examples of these longitudinal surveys are the British Household Panel Survey, the German Socio-Economic Panel, the World Values Survey and the Eurobarometer.

5.2. THE MONETARY VALUATION OF INFORMAL CARE

When informal care is valued using stated preference-based methods –namely, the contingent valuation (CV) method, conjoint analysis (CA) or discrete choice experiments (DCEs)–, it is possible to ask different groups of the population: informal carers, care recipients, the general public as a whole or only subjects who are neither carers nor care recipients (who were labeled ‘non-carers’ in Chapter 3). However, preceding studies on the monetary valuation of informal care that have used this kind of methods have focused on the carer’s standpoint (Gustavsson *et al.*, 2010; Mentzakis *et al.*, 2010) and, in a few cases, on both the carer’s and the care recipient’s perspectives (Van den Berg *et al.*, 2005; De Meijer *et al.*, 2010), but no study has obtained a monetary value for informal care from a different viewpoint. Against this background, in the study reported in Chapter 3 we elicited a monetary valuation for informal care from a sample of non-carers (excluding care recipients) and compared it with that derived from a sample of actual informal carers. The specific questions addressed in Chapter 3 are listed below:

- Is it feasible to obtain a monetary value for informal care based on the stated preferences of non-carers?
- Does this valuation differ from that elicited from a sample of informal carers?
- Is this valuation consistent with (and sensitive to) the strength of preferences over different caring tasks?
- Are carers' values for informal care in reference to a hypothetical caring situation influenced by their own experience as informal carers?

To answer these questions, we conducted a CV study involving two groups: a sample of informal carers (*n* = 202) and a sample of non-carers (*n* = 200). Respondents in both groups were asked three willingness to accept (WTA) questions in reference to a hypothetical situation: WTA for one extra hour of informal care per day in general terms ('$WTA_{general}$'); WTA for one additional hour of care per day carrying out the least preferred task ('WTA_{worst}'); and WTA for one extra hour per day undertaking the most preferred task ('WTA_{best}'). Furthermore, informal carers were enquired to state the minimum amount of money they would demand if they had to devote one additional hour per day to look after their own care recipient ('WTA_{own}').

5.2.1. Main results and discussion

From the results obtained in Chapter 3, we can conclude that it is feasible to value informal care in monetary terms from a sample of non-carers (at least using WTA). This was reflected by the small proportion of protest zeros[1] (6.5% of non-carers provided this kind of response in at least one of the three WTA questions), as well as by the low rate of inconsistent non-carers[2] (6.4%). However, there was a considerable proportion (around 25%) of insensitive (or invariant) non-carers (i.e. those who gave exactly the same value in the three hypothetical WTA questions). As a result, only a third of subjects in this sample were strictly consistent[3].

---

[1] In this study, protest zeros were defined as zero WTA values motivated by ethical objections against being compensated for taking care of a loved one.

[2] Inconsistent subjects were those who exhibited at least one of the following patterns: WTA_{best} > WTA_{worst}, $WTA_{general}$ > WTA_{worst} or WTA_{best} > $WTA_{general}$).

[3] Strictly consistent respondents were those who satisfied the ranking $WTA_{best} < WTA_{general} < WTA_{worst}$.

The results found in the sample of carers were fairly similar as those obtained in the sample of non-carers. Consequently, the distributions of carers' and non-carers' WTA values did not differ significantly. In particular, the mean WTA_{best}, $WTA_{general}$ and WTA_{worst} values were only slightly lower in the sample of carers (€5.3, €6.4 and €7.5, respectively) than in the sample of non-carers (€5.6, €6.5 and €7.9, respectively). The median WTA_{best} and WTA_{worst} values were also lower in the former group (€4.5 and €7, respectively) than in the latter (€5.5 and €9, respectively), whereas the same median $WTA_{general}$ value was obtained in both cases (€5.5).

On the other hand, the mean/median WTA_{own} values in the sample of carers (€5.2/€4.5) were even lower than the mean/median $WTA_{general}$ values obtained in the same group (€6.4/€5.5). This means that, on average, carers demanded a lower compensation for taking care of their own care recipient for one additional hour per day than if they had to devote that extra hour to look after the person described in the hypothetical scenario. This could suggest that the positive aspects associated with the provision of informal care and the presence of moral concerns played a greater role in the WTA_{own} question than in the $WTA_{general}$ question. This assumption is supported by the fact that carers were more prone to provide a protest zero in the WTA_{own} question than in any of the three hypothetical WTA questions.

On closer inspection, carers' valuations seemed to have been influenced (or 'contaminated') by their own experience providing informal care. Indeed, more than a half of them required the same compensation if they had to spend one extra hour per day taking care of the hypothetical care recipient as if they had to devote that additional time to assist their loved one. This is closely related to the fact that most carers (75%) declared they had thought of their own care recipient when they answered the hypothetical WTA questions, even though in all carers' questionnaires the $WTA_{general}$ question was posed before the WTA_{own} question. These results could also be showing that the hypothetical WTA questions were able to capture some subjective aspects associated with the provision of informal care.

It must me pointed out that our results are not directly comparable with those reported in prior WTA studies on the monetary valuation of informal care, to the extent that in our study the questions were posed in terms of the monetary compensation demanded in exchange for providing one extra hour of informal care per day, whereas other studies have used a weekly basis. With this in mind, our WTA values are lower than those reported in previous studies that have also valued informal care using the WTA method, even after adjusting for purchasing power parity (PPP). For example, from a heterogeneous sample of Dutch informal carers, Van den Berg *et al.* (2005) obtained mean/median WTA values of €10.5/€9.1 (€10/€8.7 in constant prices and PPP) per extra hour a week, which are even higher than our carers' mean/median

WTA_{worst} values (€6.6 or €6.2 in constant prices and PPP). On the other hand, the mean WTA_{own} value elicited in our sample of carers (€5.2 or €4.6 in constant prices and PPP) is close to the mean WTP value reported by Gustavsson *et al.* (2010) for a sample of Spanish informal carers in charge of patients with Alzheimer's disease (€4.7, after expressing it into a daily basis, or €4.1 in constant prices and PPP). Nevertheless, it is worth remembering that WTA and WTP values are not comparable because WTA and WTP questions use a different framing (De Meijer *et al.*, 2010).

5.2.2. Limitations and suggestions for future research

It is important to note that the aforementioned results are not generalisable and, therefore, they should be interpreted with caution. This mainly stems from two limitations of our study. First, the two samples were recruited in a nearby area (the Region of Murcia, Spain), had a rather small size and were not representative. Albeit the sample of non-carers was representative of the Spanish adult population in terms of age and sex, no representativeness was achieved in terms of socioeconomic status (education and household income). The non-representativeness of the sample of carers (with respect to the Spanish population of informal carers) was more marked because in this case we did not set up age and gender quotas. On average, compared with the Spanish population of informal carers, those in our study bore less objective burden and were younger (around 5 years). This suggests that there could be a 'self-selection' bias in the sample of carers, in the sense that most of those who agreed to participate in the study bore a mild burden. However, unlike other studies on the monetary valuation of informal care, we did not confine our attention to informal carers in charge of patients with a particular health problem (e.g. Alzheimer's disease, rheumatoid arthritis, etc.). On the other hand, care recipients were not included in the study, because we were particularly interested in examining the point of view of individuals who are unfamiliar with informal care. As a result, we could not compute a public value by aggregating the WTA values of the two samples. In this regard, we acknowledge that it would be of great interest to obtain a monetary value for informal care from a representative sample of the general population (which is comprised by carers, care recipients and others), in a similar manner in which preferences over health states are frequently elicited from a representative sample of the general public with the aim of accounting for the societal preferences.

The second major limitation of the study is that all hypothetical WTA values were elicited in reference to a specific hypothetical caring scenario (which was the same for all participants). To overcome this limitation, we propose to develop a multi-attribute descriptive system, whose dimensions would be related to the caring tasks to be performed, the time to be invested in these tasks and the health state of the care recipient. Such an instrument would enable a number of care-related situations to be characterise in terms of those attributes, as well as to derive a set of monetary values for each possible caring situation.

5.3. THE EFFECT OF LEARNING ON THE CONSISTENCY OF PREFERENCES

Although the preference reversal (PR) phenomenon is a robust and systematic failure of procedural invariance, there is some evidence that, under certain circumstances, it tends to becomes less frequent as individuals learn through the acquisition of experience with the tasks they undertake (Cox and Grether, 1996; Braga *et al.*, 2009). Hitherto, however, no study had tested the effect of learning on PR using health outcomes. In order to fill this gap, the main objective of Chapter 4 was to examine whether the frequency of PR (in relation to both health and monetary outcomes) subsides when learning is promoted. To that end, two learning effects were differentiated: the effect of the mere repetition of tasks and the combined effect of repetition and feedback on the consequences of decisions made. On this background, Chapter 4 was intended to answer the following questions:

- Is the frequency of PR attenuated as individuals acquire experience with the tasks they perform (through repetition) and learn about the consequences of their own decisions (through feedback)?
- Does learning lessen the typical discrepancy between standard preference reversal (SPR) and non-standard preference reversal (NSPR)? If so, does the standard asymmetric pattern of PR evolve towards a non-standard asymmetric pattern (with NSPR being more frequent than SPR)?
- Is PR sensitive to the nature of the outcomes used (health, money)?
- Are there gender differences in terms of PR?

To answer these questions, we conducted an experiment (with a sample of 319 undergraduate students) comprising two separate sessions, with the aim of isolating the two learning effects mentioned earlier: the effect of the repetition of tasks (first session) and the joint effect of repetition and feedback on the consequences of decisions made (second session). Preferences were obtained using two normatively equivalent methods: straight choices between two lotteries ('P-bet' and '$-bet') and separate valuations of the two lotteries elicited as certainty equivalent (CE) values. The sample was split into three groups, depending on the nature of the lottery outcomes used. Participants in Groups 2 and 3 were exposed to health outcomes ('years of life' and 'days without back pain', respectively), whereas those in Group 1 were presented with monetary outcomes.

5.3.1. Main results and discussion

We observed a strong asymmetry between SPR and NSPR throughout the experiment. On average, the former represented 50% of responses in the first session and 53% in the second one, whilst the latter accounted for 1.5% and 2.9% of answers in sessions 1 and 2, respectively. This asymmetric pattern is typically found in PR experiments and it is considered to be particularly worrying because it cannot be explained by response error alone (Cox and Grether, 1996; Braga and Starmer, 2005). The trend of responses revealed that the mere repetition of tasks did not suffice to erode SPR significantly. By contrast, the repetition of tasks in combination with feedback was more effective to reduce the presence of SPR, leading to a significant increase in the rate of consistent responses. However, the improvement in consistency was limited due to the significant upward trend in NSPR over the second session. Indeed, the results obtained at the end of the experiment were far from satisfactory. For instance, in the last round of the second session, on average, 48.1% and 4.1% of respondents incurred in SPR and NSPR, respectively, and only 38.6% of subjects were strictly consistent (i.e. they were consistent and placed a different value on each paired bet). Given that SPR and NSPR followed opposite trends over the second session, the asymmetry between the two types of PR was softened over that session, but only moderately.

The downward trend of SPR (especially over the second session) might be explained by two hypotheses: the 'discovered preference hypothesis' (Plott, 1996) and the 'constructed preference approach' (Slovic, 1995; Lichtenstein and Slovic, 2006). The former posits that individuals own a unique set of true underlying preferences, which are prior to the decisions that subjects are asked to make. The discovered preference hypothesis further assumes that preferences are not biased (i.e. they do not exhibit anomalies). However, when subjects handle unfamiliar and/or complex tasks, at first, their true preferences may be blurry and, consequently, their stated preferences may deviate from their true preferences. But if learning is fostered, for instance, by means of the repetition of tasks, feedback on the consequences of decisions made and, if possible, incentives, it is possible to lessen the discrepancy between the stated preferences and the true preferences. Therefore, the discovered preference hypothesis suggests that inexperienced stated preferences are likely to be inconsistent, but the consistency of preferences can be heightened as individuals discover their own preferences through a process of learning. This is an important point to bear in mind because preferences are usually elicited in surveys where there is neither repetition nor feedback. By contrast, the constructed preference view assumes that preferences are constructed at the time of making a decision. Thus, according to this hypothesis, there are no underlying preferences –or, if they exist, there is not a unique set of preferences (Braga and Starmer, 2005). In our view, the fall in the frequency of SPR observed in our experiment is in line with the discovered preference hypothesis –other researchers, however, could regard this trend as an instance of the construction of preferences. This pattern supports the idea that some preference anomalies (such as SPR) can be attenuated if respondents have the opportunity to learn, through repetition of tasks and feedback on the consequences of their decisions.

Conversely, the drop in the frequency of NSPR over the second session is a weird finding that could be attributable to response error and/or imprecision. It must be noted that Braga *et al.* (2009) also found the same result. What is more, in this latter study NSPR became even more prevalent than SPR by the end of the experiment, reversing the typical pattern observed in PR studies.

Regarding the differences depending on the type of outcomes used, SPR was more frequent in the two groups in which health outcomes were used (i.e. in Groups 2 and 3) than in the groups where the lotteries offered monetary outcomes (i.e. in Group 1). More specifically, the highest rates of SPR were obtained in Group 2 (where the lottery outcomes were expressed in terms of years of life in good health and death as success and failure outcomes, respectively). Since respondents in Group 2 chose the P-bet (i.e. the high probability lottery) at a significantly higher rate than those in the other two groups, our results suggest that, in the choice tasks, individuals in Group 2 were induced to be more risk averse than the rest of participants as a result of having been exposed to lotteries that entailed a risk of death. This could reinforce the preference for the P-bet over the $-bet when choosing. The substantial rates of PR in the group who dealt with years of life and death as outcomes is problematic because many health care interventions which are assessed as part of a preference elicitation exercise involve some risk of death. Future studies could investigate whether SPR is even more likely to arise when using gambles whose failure outcome is considered to be worse than death.

Given the overwhelming high rates of SPR found in Groups 2 and 3 (especially in the former), even at the end of the experiment, we consider that it is essential to test the consistency of preferences over health outcomes, given the potentially high risk of PR, and to provide separate statistics for consistent and inconsistent individuals. For example, when two or more health states are valued (e.g. using the standard gamble or the time trade-off methods), a simple consistency test is to compare the implicit ranking derived from the utility values obtained for each health state with the explicit ranking resulting from a ranking task involving those states. Likewise, it is hugely important to search for strategies to enhance the consistency of stated preferences. In this respect, against the common practice of eliciting preferences in experiments where each task is performed just once, we recommend that, whenever possible, preferences be elicited in interactive sessions, enabling respondents to learn from both the repetition of tasks and feedback on the consequences of their decisions (as in our experiment), in an attempt to minimise the risk of PR. Notwithstanding, we are aware that this solution is usually unfeasible, because it is time-consuming and expensive (Bleichrodt *et al.*, 2001). Moreover, it must be borne in mind that the repetition of tasks can induce new biases as a result of respondents' tiredness and boredom. When it is not possible to elicit preferences in interactive sessions, an alternative solution is the use of corrective formulas (e.g. based on prospect theory), able to eliminate the deviation of the elicited inconsistent preferences from the true preferences.

On the other hand, we also observed significant gender differences, insofar women were more likely to exhibit SPR than men and, in consequence, throughout the experiment the rate of consistent women was significantly lower than that of consistent men. This difference is mainly explained because women displayed a stronger risk aversion than men when choosing, as reflected by the fact that in all choices the proportion of women that selected the P-bet was significantly higher than the proportion of men that chose the same bet. This greater risk aversion of women in the choices is in line with preceding research (Hartog *et al.*, 2002; Eckel and Grossman, 2008). Furthermore, as compared with men, women tended to report lower CE values for the P-bets and higher CE values for the $-bets.

5.3.2. Limitations and suggestions for future research

As all experiments, that presented in Chapter 4 has a number of limitations that could condition our findings.

First, each session involved a relatively scarce number of rounds (i.e. three rounds for each pair of gambles). This fact could partly explain the persistence of a high frequency of SPR even at the end of the experiment. We suspect that, with more repetitions combined with feedback, the frequency of SPR would have continued moving downwards. But the main doubt is whether the presence of NSPR would have remained growing or even if, beyond a certain number of rounds, NSPR would have exceeded SPR, giving rise to a non-standard asymmetry between the two types of PR (with NSPR being more prevalent than SPR), as in the experiment undertaken by Braga *et al.* (2009). To clarify these doubts, future experiments should involve more rounds with repetition and feedback. The main problem is that, as previously mentioned, promoting learning in interactive sessions has a cost (in terms of money and time) and it can be burdensome for respondents, with the subsequent risk of inducing response error.

Furthermore, in the second session there was no repetition of choices, but only of valuation tasks (i.e. a single choice was made with each pair of lotteries). Although previous experiments have found little variation in choice behaviour (Butler and Loomes, 2007; Braga *et al.*, 2009), choices can also be sensitive to the acquisition of experience. For example, through the repetition of choices, respondents may better understand the meaning of probabilities. Therefore, it would be interesting to replicate our experiment to test whether the repetition of both valuations and choices with both repetition and feedback yields significantly different results than those we found.

Another limitation of the second session is that the way in which feedback was provided (i.e. by means of the visualisation of risk of the lotteries) only showed the risk associated with the treatments being assessed, but not the impact of those treatments on patients' quality of life. In this regard, future studies could test whether the fact of providing feedback by displaying the impact of the treatments under consideration on patients' lives (e.g. using videos or leaflets) is more effective to enhance the consistency of preferences than the mere resolution of risk of the gambles.

Moreover, it must be borne in mind that no generalisations can be drawn from our results because we used a convenience sample comprised of 319 students who were relatively homogeneous with respect to age and educational background. In consequence, we were not able to test whether age and educational level have an influence on PR. However, it is worth noting that most experiments about PR are also conducted with convenience samples, which are usually smaller in size than ours.

In the preceding lines we have put forward some suggestions for further research. Apart from them, there are other interesting issues that well deserve to be investigated in the future, such as the differences between men and women in terms of PR (exploring whether they are explained by biological and hormonal factors); the influence of age and educational level on PR;

the reproducibility of our results in other settings (e.g. using different samples and different lotteries); as well as the effect of learning on other forms of PR observed in the health economics literature –for instance, the matching-matching discrepancy (e.g. the disparity between certainty equivalent and probability equivalent values) or internal inconsistencies between variants of a same preference elicitation procedure. Furthermore, a top issue in the agenda of health economists should be to identify those methods that best capture the true underlying preferences. In this regard, qualitative surveys and techniques such as brain scanning, eye-movement, think-aloud eye-fixations, or computer cursor movements, among others, could allow us to better understand how people arrive at a decision.

REFERENCES

Abellán, J. M., Bleichrodt, H., & Pinto, J. L. (2009). The predictive validity of prospect theory versus expected utility in health utility measurement. *Jounal of Health Economics, 28*(6), 1039-1047.

Abellán, J. M., Sánchez, F. I., Martínez, J. E., & Méndez, I. (2012). Lowering the 'floor' of the SF-6D scoring algorithm using a lottery equivalent method. *Health Economics, 21*(11), 1271-1285.

Al-Janabi, H., Coast, J., & Flynn, T. N. (2008). What do people value when they provide unpaid care for an older person? A meta-ethnography with interview follow-up. *Social Science & Medicine, 67*(1), 111-121.

Al-Janabi, H., Frew, E., Brouwer, W. B. F., Rappange, D., & van Exel, N. J. A. (2010). The inclusion of positive aspects of caring in the Caregiver Strain Index: Tests of feasibility and validity. *International journal of nursing studies, 47*(8), 984-993.

Al-Windi, A. (2005). The relations between symptoms, somatic and psychiatric conditions, life satisfaction and perceived health. A primary care based study. *Health and Quality of Life Outcomes, 3,* 28.

Alden, A. L., Dale, J. A., & DeGood, D. E. (2001). Interactive effects of the affect quality and directional focus of mental imagery on pain analgesia. *Applied Psychophysiology and Biofeedbak, 26*(2), 117-126.

Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and happiness: Are Europeans and Americans different? *Journal of Public Economics, 88*(9-10), 2009-2042.

Amiel, Y., Cowell, F., Davidovitz, L., & Polovin, A. (2008). Preference reversals and the analysis of income distributions. *Social Choice and Welfare, 30*(2), 305-330.

Arno, P. S., Levine, C., & Memmott, M. M. (1999). The economic value of informal caregiving. *Health Affairs, 18*(2), 182-188.

Arrindell, W. A., Heesink, J., & Feij, J. A. (1999). The Satisfaction With Life Scale (SWLS): Appraisal with 1700 healthy young adults in The Netherlands. *Personality and Individual Differences, 26*(5), 815-826.

Arrindell, W. A., Meeuwesen, L., & Huyse, F. J. (1991). The Satisfaction With Life Scale (SWLS): Psychometric properties in a non-psychiatric medical outpatients sample. *Personality and Individual Differences, 12*(2), 117-123.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register, 58,* 4601-4614.

Arthaud-Day, M., Rode, J., Mooney, C., & Near, J. (2005). The subjective well-being construct: A test of its convergent, discriminant, and factorial validity. *Social Indicators Research, 74*(3), 445-476.

Atkinson, G., Morse-Jones, S., Mourato, S., & Provins, A. (2012). When to take “no” for an answer’? Using entreaties to reduce protests in contingent valuation studies. *Environmental and Resource Economics, 51*(4), 497-523.

Atkinson, S. E., & Halvorsen, R. (1990). The valuation of risks to life: Evidence from the market for automobiles. *The Review of Economics and Statistics, 72*(1), 133-136.

Badia, X., Roset, M., Herdman, M., & Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making, 21*(1), 7-16.

Bateman, I., Day, B., Loomes, G., & Sugden, R. (2007). Can ranking techniques elicit robust values? *Journal of Risk and Uncertainty, 34*(1), 49-66.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226-232.

Benyamini, Y., Leventhal, H., & Leventhal, E. A. E. A. (2004). Self-rated oral health as an independent predictor of self-rated general health, self-esteem and life satisfaction. *Social Science & Medicine, 59*(5), 1109-1116.

Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E., Gilson, B. S., & Morris, J. R. (1976). The sickness impact profile: Conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services, 6*(3), 393-415.

Bharmal, M., & Thomas, J. (2006). Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value in Health, 9*(4), 262-271.

Binmore, K. (1994). *Playing fair*. Cambridge, MA: MIT Press.

Binmore, K. (1999). Why experiment in economics? *Economic Journal, 109*(453), 16-24.

Blais, M. R., Vallerand, R. J., Pelletier, L. G., & Briere, N. M. (1989). L'Echelle de satisfaction de vie: Validation Canadienne-Francaise du «Satisfaction With Life Scale» (French-Canadian Validation of the Satisfaction With Life Scale). *Canadian Journal of Behavioral Science, 21*, 210-223.

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics, 88*(7-8), 1359-1386.

Bleichrodt, H. (2001). Probability weighting in choice under risk: An empirical test. *Journal of Risk and Uncertainty, 23*(2), 185-198.

Bleichrodt, H., Abellan-Perpiñan, J. M., Pinto-Prades, J. L., & Mendez-Martinez, I. (2007). Resolving inconsistencies in utility measurement under risk: Tests of generalizations of expected utility. *Management Science, 53*(3), 469-482.

Bleichrodt, H., & Johannesson, M. (1997). Characterizing QALYs by risk neutrality. *Journal of Risk and Uncertainty 15*(2), 107-114.

Bleichrodt, H., & Pinto, J. L. (2009). New evidence of preference reversals in health utility measurement. *Health Economics, 18*(6), 713-726.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science, 47*(11), 1498-1514.

Bleichrodt, H., & Quiggin, J. (2013). Capabilities as menus: A non-welfarist basis for QALY evaluation. *Journal of Health Economics, 32*(1), 128-137.

Bobinac, A., van Exel, N. J. A., Rutten, F. F. H., & Brouwer, W. B. F. (2010). Caring for and caring about: Disentangling the caregiver effect and the family effect. *Journal of Health Economics, 29*(4), 549-556.

Böckerman, P., & Ilmakunnas, P. (2009). Unemployment and self-assessed health: Evidence from panel data. *Health Economics, 18*(2), 161-179.

Böckerman, P., Johansson, E., & Saarni, S. I. (2011). Do established health-related quality-of-life measures adequately capture the impact of chronic conditions on subjective well-being? *Health Policy, 100*(1), 91-95.

Braga, J., Humphrey, S. J., & Starmer, C. (2009). Market experience elimnates some anomalies—and creates new ones. *European Economic Review, 53*(4), 401-416.

Braga, J., & Starmer, C. (2005). Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics, 32*(1), 55-89.

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment, 3*(9), i-iv, 1-164.

Brazier, J., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., *et al.* (1992). Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *BMJ, 305*(6846), 160-164.

Brazier, J., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*(9), 851-859.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*(2), 271-292.

Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ, 13*(9), 873-884.

Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology, 51*(11), 1115-1128.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 36*(8), 917-927.

Bromley, D. W. (1995). Property rights and natural resource damage assessments. *Ecological Economics, 14*(2), 129-135.

Brouwer, W. B. F., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. H. (2008). Welfarism vs. extrawelfarism. *Journal of Health Economics, 27*(2), 325-338.

Brouwer, W. B. F., Van Exel, N. J. A., Van den Berg, B., Van den Bos, G. A. M., & Koopmanschap, M. A. (2005). Process utility from providing informal care: The benefit of caring. *Health Policy, 74*(1), 85-99.

Brouwer, W. B. F., van Exel, N. J. A., van Gorp, B., & Redekop, W. K. (2006). The CarerQol instrument: A new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Quality of Life Research, 15*(6), 1005-1021.

Brown, T. C., & Gregory, R. (1999). Why the WTA - WTP disparity matters. *Ecological Economics, 28*(3), 323-335.

Buckingham, K., & Devlin, N. (2006). A theoretical framework for TTO valuations of health. *Health Economics, 15*(1), 1149-1154.

Buckner, L., Yeandle, S., University of Leeds, & Carers UK (2011). Valuing carers 2011. Calculating the value of carers' support

Butler, D. J., & Loomes, G. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review, 97*(1), 277-297.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial ncentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1-3), 7-42.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfaction*. New York: Russell.

Canadian Agency for Drugs and Technologies in Health (2006). *Guidelines for the economic evaluation of health technologies* (3rd ed.). Ottawa: Canadian Agency for Drugs and Technologies in Health.

Carson, R. T. (1991). Constructed markets. In J. B. Braden, & C. D. Kolstad (Eds.), *Measuring the demand for environmental quality*. Amsterdam: North Holland.

Clarke, P. M. (1998). Cost–benefit analysis and mammographic screening: a travel cost approach. *Journal of Health Economics, 17*(6), 767-787.

Claxton, K., Walker, S., Palmer, S., & Sculpher, M. (2010). Appropriate perspectives for health care decisions. CHE Research Paper 54.

Coe, N. B., & Van Houtven, C. H. (2009). Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health Economics, 18*(9), 991-1010.

Cohen, C. A., Colantonio, A., & Vernich, L. (2002). Positive aspects of caregiving: Rounding out the caregiver experience. *International Journal of Geriatric Psychiatry, 17*(2), 184-188.

Comunidad Autónoma de la Región de Murcia (2011). BORM nº.151. Orden de la Consejería de Política Social, Mujer e Inmigración, por la que se establecen los precios públicos por la prestación de los servicios del Sistema de Autonomía y Atención a la Dependencia y del Sistema de Servicios Sociales con financiación Pública en la Comunidad Autónoma de la Región de Murcia.

Cookson, R. (2003). Willingness to pay methods in health care: A sceptical view. *Health Economics, 12*(11), 891-894.

Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics, 17*(1), 13-35.

Corporation for National and Community Service, O. o. R. a. P. D. (2007). The health benefits of volunteering. A review of recent research. Washington, DC.

Corso, P. S., Hammitt, J. K., Graham, J. D., Dicker, R. C., & Goldie, S. J. (2002). Assessing preferences for prevention versus treatment using willingness to pay. *Medical Decision Making, 22*(5 Suppl), S92-101.

Coursey, D., Hovis, J., & Schultz, W. (1987). The disparity between willingness to accept and willingness to pay measures of value. *Quartely Journal of Economics, 102*(3), 679-690.

Cox, J. C., & Grether, D. M. (1996). The preference reversal phenomenon: Response mode, markets and incentives. *Economic Theory, 7*(3), 381-405.

Cubí-Mollá, P., & Herrero, C. (2012). Quality of life lost due to non-fatal road traffic injuries. *Health Economics, 21*(5), 528-550.

Cubitt, R. P., Munro, A., & Starmer, C. (2004). Testing explanations of preference reversal. *The Economic Journal, 114*(497), 709-726.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics, 1*(2), 115-131.

Culyer, A. J. (1989). The normative economics of health care finance and provision. *Oxford Review of Economic Policy, 5*(1), 34-58.

Cummings, R. G., Brookshire, D. S., & Schulze, W. D. (1986). *Valuing environmental goods: A state of the arts assessment of the contingent valuation method*. Totowa, NJ: Roweman and Allanheld.

Chu, Y.-P., & Chu, R.-L. (1990). The subsidence of preference reversals in simplified and marketlike experimental settings: A note. *The American Economic Review, 80*(4), 902-911.

De Meijer, C., Brouwer, W. B. F., Koopmanschap, M. A., Van den Berg, B., & Van Exel, N. J. A. (2010). The value of informal care – A further investigation of the feasibility of contingent valuation in informal caregivers. *Health Economics, 19*(7), 755-771.

Delquié, P. (1993). Inconsistent trade-offs between attributes: New evidence in preference assessment biases. *Management Science, 39*(11), 1382-1395.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*(2), 197-229.

Devlin, N. J., Tsuchiya, A., Buckingham, K., & Tilling, C. (2011). A uniform time trade off method for states better and worse than dead: Feasibility study of the 'lead time' approach. *Health Economics, 20*(3), 348-361.

Di Tella, R., MacCulloch, R., & Oswald, A. J. (2001). Preferences over inflation and unemployment. Evidence from surveys of happiness. *American Economic Review, 91*(1), 335-341.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542-575.

Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Journal of Happiness Studies, 7*(4), 397-404.

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being, 3*(1), 1-43.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71-75.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (2006). Understanding scores on the Satisfaction With Life Scale.

Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*(1), 1-31.

Dolan, P. (1997). Modelling valuations for EuroQol health states. *Medical Care, 35*(11), 1095-1108.

Dolan, P. (2000). The measurement of health-related quality of life for use in resource allocation decisions in health care. In A. J. Culyer, & J. P. Newhouse (Eds.), *Handbook of Health Economics* (1st ed., Vol. 1, pp. 1724-1760). Amsterdam: Elsevier

Dolan, P. (2008a). Developing methods that really do value the 'Q' in the QALY. *Health Economics, Policy and Law, 3*(1), 69-77.

Dolan, P. (2008b). Thinking about it: Thoughts about health and valuing QALYs. Imperial College, London, Imperial College Business School.

Dolan, P. (2011). Using happiness to value health. London: Office of Health Economics.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). The time trade-off: Results from a general population study. *Health Economics, 5*(2), 141-154.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *Economic Journal, 118*(525), 215-234.

Dolan, P., & Metcalfe, R. (2012). Valuing health: A brief report on subjective well-being versus preferences. *Medical Decision Making, 32*(4), 578-582.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology, 29*(1), 94-122.

Dolan, P., & White, M. P. (2007). How can measures of subjective well-being be used to inform public policy? *Perspectives on Psychological Science, 2*(1), 71-85.

Donaldson, C. (1999). Developing the method of 'willingness to pay' for assessment of community preferences for health care ('EuroWill'). Final Report to the European Commission Biomed 2 Programme.

Donaldson, C., Mason, H., & Shackley, P. (2012). Contingent valuation in health care. In A. M. Jones (Ed.), *The Elgar Companion to Health Economics* (2nd ed., pp. 425-435). Cheltenham, UK: Edward Elgar Publishing.

Drichoutis, A., Nayga, R., & Klonaris, S. (2010). The effects of induced mood on preference reversals and bidding behavior in experimental auctions. MPRA Paper, Nº 31085.

Drummond, M., Brixner, D., Gold, M. R., Kind, P., McGuire, A., & Nord, E. (2009). Toward a consensus on the QALY. *Value Health, 12 Suppl 1*, S31-35.

Drummond, M., Sculpher, M., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programmes* (3rd ed.). Oxford: Oxford University Press.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1-17.

Edgeworth, F. (1881/1967). *Mathematical Psychics*. New York: Kelley.

European Commision (2011). Eurobarometer Survey. Available at: http://ec.europa.eu/public_opinion/index_en.htm.

EuroQol Group (1990). EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199-208.

Falk, D. (2005). Brain lateralization in primates and its evolution in hominids. *American Journal of Physical Anthropology, 30*(S8), 107-125.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., *et al.* (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care, 40*(2), 113-128.

Ferraro, K. F., & Albrecht-Jensen, C. M. (1991). Does religion influence adult health? *Journal for the Scientific Study of Religion, 30*(2), 193-202.

Ferreira, L., Ferreira, P., Pereira, L., & Brazier, J. (2008). An application of the SF-6D to create heath values in Portuguese working age adults. *Journal of Medical Economics*, *11*(2), 215-233.

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? . *Economic Journal, 114*(497), 641-659.

Ferrer-i-Carbonell, A., & Van Praag, B. M. (2002). The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. *Health Economics, 11*(8), 709-722.

Fischer, G. W., Carmon, Z., Ariely, D., & Zauberman, G. (1999). Goal-based construction of preferences: Task goals and the prominence effect. *Management Science, 45*(8), 1057-1075.

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology, 19*(3), 580-597.

Franks, P., Gold, M. R., & Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social Science & Medicine, 56*(12), 2505-2514.

Fraser, G. E. (1999). Associations between diet and cancer, ischemic heart disease, and all-cause mortality in non-Hispanic white California Seventh-day Adventists. *The American Journal of Clinical Nutrition, 70*(3 Suppl), 532S-538S.

Frederikse, M., Aylward, E., Barta, P., & Pearlson, G. (1999). Sex differences in the inferior parietal lobule. *Cerebral Cortex, 9*(8), 896-901.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature, 40*(2), 402-435.

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. *Journal of Personality and Social Psychology, 61*(3), 427-434.

Gandjour, A. (2001). Is subjective well-being a useful parameter for allocating resources among public interventions? *Health Care Anal, 9*(4), 437-447.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across Europe. *Journal of Health Economics, 30*(1), 200-213.

Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal, 324*(7351), 1417.

Garrido, S., Méndez, I., & Abellán, J. M. (2013). Analysing the simultaneous relationship between life satisfaction and health-related quality of life. *Journal of Happiness Studies, 14*(3), 1813-1838.

George, L. K., & Landerman, R. (1984). Health and subjective well-being: A replicated secondary data analysis. *International Journal of Aging & Human Development, 19*(2), 133-156.

Gerdtham, U.-G., & Johannesson, M. (2001). The relationship between happiness, health and socio-economic factors: Results based on Swedish micro data. *Journal of Socio-Economics, 30*(1), 553-557.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review, 94*(2), 236-254.

Goodrich, K., Kaambwa, B., & Al-Janabi, H. (2012). The inclusion of informal care in applied economic evaluation: A review. *Value in Health, 15*(6), 975-981.

Graham, C., Higuera, L., & Lora, E. (2011). Which health conditions cause the most unhappiness? *Health Economics, 20*(12), 1431-1447.

Grandy, S., & Fox, K. M. (2012). Change in health status (EQ-5D) over 5 years among individuals with and without type 2 diabetes mellitus in the SHIELD longitudinal study. *Health and Quality of Life Outcomes, 10*(1), 99.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review, 69*(4), 623-638.

Grieve, R., Grishchenko, M., & Cairns, J. (2009). SF-6D versus EQ-5D: Reasons for differences in utility scores and impact on reported cost-utility. *The European Journal of Health Economics, 10*(1), 15-23.

Grossman, M., & Wendy, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology, 65*(5), 1010-1022.

Grundmann, E. (1992). Cancer morbidity and mortality in USA Mormons and Seventh-day Adventists. *Archives of Anatomy Citology and Pathology, 40*(2-3), 73-78.

Gustavsson, A., Jönsson, L., McShane, R., Boada, M., Wimo, A., & Zbrozek, A. S. (2010). Willingness-to-pay for reductions in care need: Estimating the value of informal care in Alzheimer's disease. *International Journal of Geriatric Psychiatry, 25*(6), 622-632.

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine, 118*(8), 622-629.

Gyldmark, M., & Morrison, G. C. (2001). Demand for health care in Denmark: Results of a national sample survey using contingent valuation. *Social Science & Medicine, 53*(8), 1023-1036.

Haab, T. C., & McConnell, K. E. (2002). *Valuing environmental and natural resources: The econometrics of non-market valuation* (1st ed.). Cheltenham, UK: Edward Elgar.

Halstead, J. M., Luloff, A. E., & Stevens, T. H. (1992). Protest bidders in contingent valuation. *Northeastern Journal of Agricultural and Resource Economics, 91*(2), 160-169.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking measured risk aversion to individual characteristics. *Kyklos, 55*(1), 3-26.

Hassink, W. H. J., & Van den Berg, B. (2011). Time-bound opportunity costs of informal care: Consequences for access to professional care, caregiver support, and labour supply estimates. *Social Science & Medicine, 73*(10), 1508-1516.

Hastrup, L. H., vVn Den Berg, B., & Gyrd-Hansen, D. (2011). Do informal caregivers in mental illness feel more burdened? A comparative study of mental versus somatic illnesses. *Scandinavian Journal of Public Health, 39*(6), 598-607.

Heady, B., Schupp, J., Tucci, I., & Wagner, G. G. (2008). Authentic happiness theory supported by impact of religion on life satisfaction: A longitudinal analysis with data for Germany. *IZA Discussion Paper No. 3915*.

Health Information and Quality Authority (2010). *Guidelines for the economic evaluation of health technologies in Ireland*: Health Information and Quality Authority.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling, 20*(2), 331-360.

Helliwell, J. F. (2011). How can subjective well-being be improved? Papers in honour of Ian Stewart. In F. Gorbet, & A. Sharpe (Eds.), *New directions for intelligent government in Canada*. Ottawa, Ontario: Centre for the Study of Living Standards.

Helliwell, J. F., & Putnam, R. (2004). The social context of well-being. *Philosophical Transactions of the Royal Society London, 359*, 1435-1446.

Hepburn, C. G., & Barling, J. (1996). Eldercare responsibilities, interrole conflict, and employee absence: A daily study. *Journal of Occupational Health Psychology, 1*(3), 311-318.

Hershey, J. C., & Schoemaker, P. J. H. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science, 31*(10), 1213-1231.

Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics, 8*(4), 325-345.

Hirst, M. (2005). Carer distress: A prospective, population-based study. *Social Science & Medicine, 61*(3), 697-708.

Hoefman, R. J., van Exel, N. J. A., de Jong, S. L., Redekop, W. K., & Brouwer, W. B. F. (2011). A new test of the construct validity of the CarerQol instrument: Measuring the impact of informal care giving *Quality of Life Research, 20*(6), 875-887.

Hoefman, R. J., van Exel, N. J. A., Rose, J. M., van de Wetering, E. J., & Brouwer, W. B. F. (2013). A discrete choice experiment to obtain a tariff for valuing informal care situations measured with the CarerQol instrument. *Medical Decision Making*.

Hughes, S. L., Giobbie-Hurder, A., Weaver, F. M., Kubal, J. D., & Henderson, W. (1999). Relationship between caregiver burden and health-related quality of life. *The Gerontologist, 39*(5), 534-545.

Hultell, D., & Petter Gustavsson, J. (2008). A psychometric evaluation of the Satisfaction with Life Scale in a Swedish nationwide sample of university students. *Personality and Individual Differences, 44*(5), 1070-1079.

Hunt, S., McEwen, J., & McKenna, S. (1985). Measuring health status: a new tool for clinicians and epidemiologists. *The Journal of the Royal College of General Practitioners, 35*(273), 185-188.

Hurley, J. (2000). An overview of the normative economics of the health care sector. In A. J. Culyer, & J. P. Newhouse (Eds.), *Handbook of Health Economics* (1st ed., Vol. 1, Part A, pp. 55-118). Amsterdam: Elsevier

IMSERSO (2004). Survey on Informal Care 2004. Available at: http://envejecimiento.csic.es/estadisticas/encuestas/index.html.

Johannesson, M., & Johansson, P.-O. (1997). Is the valuation of a QALY gained independent of age? Some empirical evidence. *Journal of Health Economics, 16*(5), 589-599.

Johnston, K., Buxton, M. J., Jones, D. R., & Fitzpatrick, R. (2001). Collecting resource use data for costing in clinical trial. In A. Stevens, K. Abrams, J. Brazier, R. Fitzpatrick, & R. Lilford (Eds.), *The advanced handbook of methods in evidencebased healthcare* (pp. 215-230). London: SAGE Publication.

Juniper, E. F., Guyatt, G. H., Epstein, R. S., Ferrie, P. J., Jaeschke, R., & Hiller, T. K. (1992). Evaluation of impairment of health related quality of life in asthma: Development of a questionnaire for use in clinical trials. *Thorax, 47*(2), 76-83.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325-1348.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being *Journal of Economic Perspectives, 20*(1), 3-24.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004). The Day Reconstruction Method (DRM): Instrument Documentation.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics, 112*(2), 375-405.

Kalsekar, I., Wagner, J. S., Dibonaventura, M., Bates, J., Forbes, R., & Hebden, T. (2012). Comparison of health-related quality of life among patients using atypical antipsychotics for treatment of depression: Results from the National Health and Wellness Survey. *Health and Quality of Life Outcomes, 10*, 81.

Kaplan, R. M., Ganiats, T. G., Sieber, W. J., & Anderson, J. P. (1998). The quality of well-being scale: Critical similarities and differences with SF-36. *International Journal for Quality in Health Care, 10*(6), 509-520.

Keeney, R. L., & Raiffa, H. (1993). Unidimensional utility theory. In R. L. Keeney, & H. Raiffa (Eds.), *Decision with multiple objectives* (pp. 131-218). Cambridge: Cambridge University Press.

Kendig, H., Browning, C. J., & Young, A. E. (2000). Impacts of illness and disability on the well-being of older people. *Disability & Rehabilitation, 22*(1-2), 15-22.

Kim-Prieto, C., Diener, E., Tamir, M., Scollon, C., & Diener, M. (2005). Integrating the diverse definitions of happiness: A time-sequential framework of subjective well-being. *J Happiness Stud, 6*(3), 261-300.

Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. *British Medical Journal, 316*(7133), 736-741.

Kind, P., Lafata, J. E., Matuszewski, K., & Raisch, D. (2009). The use of QALYs in clinical and patient decision-making: Issues and prospects. *Value in Health, 12*(Suppl I), S27-S30.

Kirkcaldy, B., & Furnham, A. (2000). Positive affectivity, psychological well-being, accident- and traffic-deaths and suicide: An international comparison. *Studia Psychologica, 42*(1-2), 97-104.

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics, 99*(3), 507-521.

Knetsch, J. L., Tang, F.-F., & Thaler, R. H. (2001). The endowment effect and repeated market trials: Is the Vickrey auction demand revealing? *Experimental Economics, 4*(3), 257-269.

Koivumaa-Honkanen, H., Honkanen, R., Koskenvuo, M., Viinamäki, H., & Kaprio, J. (2002). Life dissatisfaction as a predictor of fatal injury in a 20-year follow-up. *Acta Psychiatrica Scandinavica, 105*(6), 444-450.

Koivumaa-Honkanen, H., Koskenvuo, M., Honkanen, R. J., Viinamäki, H., Heikkila, K., & Kaprio, J. (2004). Life dissatisfaction and subsequent work disability in an 11-year follow-up. *Psychological Medicine, 34*(2), 221-228.

Koopmanschap, M. A., Van Exel, N. J. A., Van den Berg, B., & Brouwer, W. B. F. (2008). An overview of methods and applications to value informal care in economic evaluations of healthcare. *Pharmacoeconomics, 26*(4), 269-280.

Kwan, K. Y. (2010). Life satisfaction and self-assessed health among adolescents in Hong Kong. *Journal of Happiness Studies, 11*, 383-393.

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., & Burgess, L. (2013). Best worst discrete choice experiments in health: Methods and an application. *Social Science & Medicine, 76*(1), 74-82.

Layard, R. (2005). *Happiness: Lessons from a new science*. New York: Penguin Press.

Lefcourt, H. M., Davidson-Katz, K., & Kuenemen, K. (1990). Humor and immune system functioning. *Humor, 3*(3), 305-321.

Leppo, K., Ollila, E., Peña, S., Wismar, M., & Cook, S. (2013). Health in all policies: Seizing opportunities, implementing policies. Ministry of Social Affairs and Health, Finland.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: State of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46-55.

Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas *Journal of Experimental Psychology, 101*(1), 16-20.

Lichtenstein, S., & Slovic, P. (2006). The construction of preference: An overview. In S. Lichtenstein, & P. Slovic (Eds.), *The construction of preference* (pp. 1-40). Cambridge ; New York Cambridge University Press.

Lilly, M. B., Laporte, A., & Coyte, P. C. (2007). Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. *Milbank Quarterly, 85*(4), 641-690.

Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology, 89*, 390-397.

Liu, J. L. Y., Maniadakis, N., Gray, A., & Rayner, M. (2002). The economic burden of coronary heart disease in the UK *Heart, 88*(6), 597-603.

Longworth, L., & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics, 12*(12), 1061-1067.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics, 8*(4), 301-323.

Loomes, G., Starmer, C., & Sugden, R. (2003). Do anomalies disappear in repeated markets? *The Economic Journal, 113*(486), C153-C166.

Loomes, G., Starmer, C., & Sugden, R. (2010). Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets. *Journal of Economic Psychology, 31*(3), 374-387.

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling, 3*(3), 57-72.

Lubetkin, E. I., Jia, H., Franks, P., & Gold, M. R. (2005). Relationship among sociodemographic factors, clinical conditions, and health-related quality of life: Examining the EQ-5D in the U.S. general population. *Quality of Life Research, 14*(10), 2187-2196.

Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71*(3), 616-628.

Luce, B. R., Wanning, W. G., Siegel, J. E., & Lipscomb, J. (1996). Estimating costs in cost-effectiveness analysis. In M. R. Gold, J. E. Siegel, L. B. Russell, & M. C. Weinstein (Eds.), *Cost-effectiveness in health and medicine* (pp. 176-213). New York: Oxford University Press.

Lundeberg, M. A., Fox, P. W., & Punccohar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*(1), 114-121.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131*(6), 803-855.

Llewellyn-Thomas, H. A., Sutherland, H. J., Tibshirani, R., Ciampi, A., Till, J. E., & Boyd, N. F. (1982). The measurement of patients' values in medicine. *Medical Decision Making, 2*(4), 449-462.

Macran, S., Weatherly, H., & Kind, P. (2003). Measuring population health: A comparison of three generic health status measures. *Medical Care, 41*(2), 218-231.

Mahoney, F. I., & Barthel, D. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal, 14*(2), 56-61.

Maier, H., & Smith, J. (1999). Psychological predictors of mortality in old age. *The Journals of Gerontology Series B: Psychological Sciences & Social Sciences, 54B*(1), 44-54.

Marra, C., Woolcott, J., Kopec, J., Shojania, K., Offer, R., Brazier, J., *et al.* (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Social Science & Medicine, 60*(7), 1571-1582.

Martínez-Granado, M., & Ruiz-Castillo, J. (2002). The decisions of Spanish youth: A cross-section study. *Journal of Population Economics, 15*(2), 305-330.

McDaid, D. (2001). Estimating the costs of informal care for people with Alzheimer's disease: methodological and practical challenges. *International Journal of Geriatric Psychiatry, 16*(4), 400-405, doi:10.1002/gps.353.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). New York: Academic Press.

Mentzakis, E., Ryan, M., & McNamee, P. (2010). Using discrete choice experiments to value informal care tasks: Exploring preference heterogeneity. *Health Economics, 20*(8), 930-944.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: A preliminary normative analysis. *Social Science & Medicine, 55*(12), 2149-2158.

Miller, L., Davies, M., & Greenwald, S. (2000). Religiosity and substance use and abuse among adolescents in the National Comorbidity Survey. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(9), 1190-1197.

Mishan, E. J., & Quah, E. (2007). *Cost-benefit analysis* (5th ed.). New York Routledge.

Molloy, G. J., Johnston, D. W., Johnston, M., Gao, C., Witham, M. D., Struthers, A. D., *et al.* (2008). Using the demand-control model of job strain to predict caregiver burden and caregiver satisfaction in the informal caregivers of heart failure patients. *British Journal of Health Psychology, 13*(3), 401-417.

Montgomery, R. J. V., Gonyea, J. G., & Hooyman, N. R. (1985). Caregiving and the experience of subjective and objective burden. *Family Relations, 34*(1), 19-26.

Mukuria, C., & Brazier, J. (2013). Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. *Social Science & Medicine, 77*, 97-105.

National Institute for Health and Clinical Excellence (2007). *Donepezil, galantamine, rivastigmine (review) and memantine for the treatment of Alzheimer's disease (amended September 2007)*. London: National Institute for Health and Clinical Excellence.

National Institute for Health and Clinical Excellence (2013). *Guide to the methods of technology appraisal* London: National Institute for Health and Clinical Excellence.

National Statistics Institute (2006). National Health Survey 2006. Available at: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t15/p419&file=inebase&L=1.

National Statistics Institute (2009). Living Conditions Survey 2009. Available at: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t25/p453&file=inebase.

Norton, E. C. (2000). *Long-term care* (Handbook of Health Economics). Amsterdam: Elsevier.

Okulicz-Kozaryn, A. (2009). Religiosity and life satisfaction across nations. *Mental Health, Religion & Culture, 13*(2), 155-169.

Okun, M. A., & George, L. K. (1984). Physician- and self-ratings of health, neuroticism and subjective well-being among men and women. *Personality and Individual Differences, 5*(5), 533-539.

Oliva, J., & Osuna, R. (2009). Los costes de los cuidados informales en España. *Presupuesto y Gasto Público, 56*(3), 163-181.

Oliva, J., Osuna, R., & Jorgensen, N. (2007). Los costes de los cuidados informales en España. In S. Jiménez-Martín (Ed.), *Aspectos económicos de la dependencia y el cuidado informal en*

España (pp. 177-214): Universidad Pompeu Fabra, Fundación de Estudios de Economía Aplicada, Fundación AstraZeneca.

Oliver, A. (2003). The internal consistency of the standard gamble: Tests after adjusting for prospect theory. *Journal of Health Economics, 22*(4), 659-674.

Oliver, A. (2006). Further evidence of preference reversals: Choice, valuation and ranking over distributions of life expectancy. *Journal of Health Economics, 25*(5), 803-820.

Oliver, A. (2013a). Testing procedural invariance in the context of health. *Health Economics, 22*(3), 272-288.

Oliver, A. (2013b). Testing the rate of preference reversal, in personal and social decision-making. *Journal of Health Economics*, doi:http://dx.doi.org/10.1016/j.jhealeco.2013.10.003.

Oliver, A., & Sorenson, C. (2008). Importance of preference reversals in the valuation of health and healthcare. *Expert Review of Pharmacoeconomics & Outcomes Research, 8*(2), 95-99.

Olsen, J. A., & Donaldson, C. (1998). Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine, 46*(1), 1-12.

Olsen, J. A., Donaldson, C., & Shackley, P. (2005). Implicit versus explicit ranking: On inferring ordinal preferences for health care programmes based on differences in willingness-to-pay. *Journal of Health Economics, 24*(5), 990-996.

Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics, 92*(5–6), 1061-1077.

padash, Z., Denavi, S. R., & botlani, S. (2012). The study of efficacy of cognitive therapy basis on positive psychology on subjective well-being. *International Journal of Business and Social Science, 3*(10).

Parkin, D., & Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics, 15*(7), 653-664.

Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. In E. Diener (Ed.), *Assessing Well-Being* (Vol. 39, pp. 101-117, Social Indicators Research Series): Springer Netherlands.

Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction With Life Scale: Evidence for the cross-method convergence of well-being. *Journal of Personality Assessment, 57*, 149-161.

Pearlin, L. I., Mullan, J. T., Semple, S. J., & Skaff, M. M. (1990). Caregiving and the stress process: an overview of concepts and their measures. *Gerontologist, 30*(5), 583-594.

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: the full life versus the empty life. *Journal of Happiness Studies, 6*(1), 25-41.

Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics, 14*(11), 1169-1189.

Pharmaceutical Benefits Advisory Committe (2008). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committeee (Version 4.3.)* Australian Government. Department of Health and Ageing.

Pinto, J. L., & Abellán, J. M. (2005). Measuring the health of populations: The veil of ignorance approach. *Health Economics, 14*(1), 69-82.

Pinto, J. L., Rodríguez, E., Castells, X., Gracia, X., & Sánchez, F. I. (2000). *El establecimiento de prioridades en cirugía electiva* Ministerio de Sanidad y Consumo.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28*(1), 206-254.

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: The discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perleman, & C. Schmidt (Eds.), *Rational foundations of economic behavior* (pp. 225-250). London and New York: Macmillan and St. Martin's.

Posnett, J., & Jan, S. (1996). Indirect cost in economic evaluation: The opportunity cost of unpaid inputs. *Health Economics, 5*(1), 13-23.

Powdthavee, N., & Van den Berg, B. (2011). Putting different price tags on the same health condition: Re-evaluating the well-being valuation approach. *Journal of Health Economics, 30*(5), 1032-1043.

Puig-Junoy, J., Sáez, M., & Martínez, E. (1998). Why do patients prefer hospital emergency visits? A nested multinomial logit analysis for patient-initiated contacts. *Health Care Management Science, 1*(1), 39-52.

Richardson, J., Olsen, J. A., Hawthorne, G., Mortimer, D., & Smith, R. (1998). The measurement and valuation of quality of life in economic evaluation. Working paper 97.

Robinson, A., Dolan, P., & Williams, A. (1997). Valuing health status using VAS and TTO: What lies behind the numbers? *Social Science & Medicine, 45*(8), 1289-1297.

Robinson, A., & Spencer, A. (2006). Exploring challenges to TTO utilities: Valuing states worse than dead. *Health Economics, 15*(4), 393-402.

Robinson, A., Thomson, R., Parkin, D., Sudlow, M., & Eccles, M. (2001). How patients with atrial fibrillation value different health outcomes: a standard gamble study. *Journal of Health Services Research & Policy, 6*(2), 92-98.

Rodríguez, E. (2012). La medición de la dependencia: Qué tenemos y qué necesitamos. *Cátedra de Economía de la Salud, Universidad Pablo de Olavide.*

Rodríguez, E., & Pinto, J. L. (2000). The social value of health programmes: Is age a relevant factor? *Health Economics, 9*(7), 611-621.

Russell, L. B. (1992). Opportunity costs in modern medicine. *Health Affairs, 11*(2), 162-169.

Russell, L. B., Gold, M. R., Siegel, J. E., Daniels, N., & Weinstein, M. C. (1996a). The role of cost-effectiveness analysis in health and medicine. *The Journal of the American Medical Association, 276*(14), 1172-1177.

Russell, L. B., Siegel, J., Daniels, N., Gold, M. R., Luce, B. R., & Mandelblatt, J. S. (1996b). Cost-effectiveness analysis as a guide to resource allocation in health: Roles and limitations. In M. R. Gold, J. E. Siegel, L. B. Russell, & M. C. Weinstein (Eds.), *Cost-effectiveness in health and medicine* (pp. 3-24). New York: Oxford University Press.

Rutten-van Mölken, M. P. M. H., Bakker, C. H., Van Doorslaer, E. K. A., & Van der Linden, S. (1995). Methodological issues of patient utility measurement. Experience from two clinical trials. *Medical Care, 33*(9), 922-937.

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2010). *Using discrete choice experiments to value health and health care* (The economics of non-market goods and resources). Doordrecht, The Netherlands: Springer.

Sabes-Figuera, R., McCrone, P., Hurley, M., King, M., Donaldson, A. N., & Ridsdale, L. (2010). The hidden cost of chronic fatigue to patients and their families. *BMC Health Services Research, 10*(56).

Schkade, D., & Johnson, E. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes, 44*(2), 203-231.

Schubert, R., Brown, M., Gysler, M., & Wolfgang, H. (1999). Financial decision-making: Are women really more risk-averse? *The American Economic Review, 89*(2), 381-385.

Schulz, R., Newsom, J., Mittelmark, M., Burton, L., Hirsch, C., & Jackson, S. (1997). Health effects of caregiving: The caregiver health effects study: An ancillary study of the Cardiovascular Health Study. *Annals of Behavioral Medicine*, *19*(2), 110-116.

Schwappach, D. L. B. (2003). The equivalence of numbers: The social value of avoiding health decline: An experimental web-based study. *BMC Medical Informatics & Decision Making, 2,* 3.

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys, 16,* 621-655.

Sen, A. (1993). Capability and well-being. In M. Nussbaum, & A. Sen (Eds.), *The quality of life* (pp. 30-54). Oxford: Clarendon Press.

Shackley, P., & Donaldson, C. (2002). Should we use willingness to pay to elicit community preferences for health care?: New evidence from using a ‘marginal’ approach. *Journal of Health Economics, 21*(6), 971-991.

Shields, M. A., & Price, S. W. (2005). Exploring the economic and social determinants of psychological well-being and perceived social support in England. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 168*(3), 513-537.

Shogren, J. F., Cho, S., Koo, C., List, J., Parch, C., Polo, P., *et al.* (2001). Auction mechanisms and the measurement of WTP and WTA. *Resource and Energy Economics, 23*(2), 97-109.

Siahpush, M., Spittal, M., & Singh, K. (2008). Happiness and life satisfaction prospectively predict self-rated health, physical health, and the presence of limiting, long-term health conditions. *American Journal of Health Promotion*, *23*(1), 18-26.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology Human Perception and Performance, 1*(3), 280-287.

Slovic, P. (1995). The construction of preferences. *American Psychologist, 50*(5), 364-371.

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R. Hogarth (Ed.), *Insights in Decision-Making* (pp. 5-27). Chicago: University of Chicago Press.

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology, 78*(3, Part 2), 1-18.

Smith, T. W., Ruiz, J. M., & Uchino, B. N. (2004). Mental activation of supportive ties, hostility, and cardiovascular reactivity to laboratory stress in young men and women. *Health Psychology, 23*(5), 476-485.

Smith, V., & Walker, J. (1993). Monetary rewards and decision costs in experimental economic. *Economic Inquiry, 31*(2), 245-261.

Spanish Centre for Sociological Research (2001). Latinobarometer 2001. Available at: http://www.cis.es/cis/opencms/EN/index.html.

Stalmeier, P. F., & Bezembinder, T. G. (1999). The discrepancy between risky and riskless utilities: A matter of framing? . *Medical Decision Making, 19*(4), 435-447.

Stalmeier, P. F. M., Chapman, G. B., de Boer, A. G. E. M., & van Lanschot, J. J. B. (2001). A fallacy of the multiplicative QALY model for low-quality weights in students and parents judging hypothetical health states. *International Journal of Technology Assessment in Health Care, 17*(04), 488-496.

Stalmeier, P. F. M., & Verheijen, A. L. (2012). Maximal endurable time states and the standard gamble: More preference reversals. *The European Journal of Health Economics*, *14*(6), 1-7.

Stalmeier, P. F. M., & Wakker, P. P. (1997). Preference reversals: Violations of unidimensional procedure invariance. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1196-1205.

Stamuli, E. (2011). Health outcomes in economic evaluation: Who should value health? *British Medical Bulletin, 97*(1), 197-210.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332-382.

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review*, *81*(4), 971-978.

Stiggelbout, A. M. (2006). Health state classification systems: How comparable are our cost-effectiveness ratios? *Medical Decision Making*, *26*(3), 223-225.

Stiglitz, J., Sen, A., & Fitoussi, J. P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. Available at: www.stiglitz-sen-fitoussi.fr [24 May 2011].

Strazzera, E., Genius, M., Scarpa, R., & Hutchinson, G. (2003). The effect of protest votes on the estimates of WTP for use values of recreational sites. *Environmental and Resource Economics, 25*(4), 461-476.

Strine, T., Chapman, D., Balluz, L., Moriarty, D., & Mokdad, A. (2008). The associations between life satisfaction and health-related quality of life, chronic illness, and health behaviors among U.S. community-dwelling adults. *Journal of Community Health, 33*(1), 40-50.

Sumner, W., & Nease, R. F. (2001). Choice-matching preference reversals in health outcome assessments. *Medical Decision Making*, *21*(3), 208-218.

Sutherland, H. J., Llewellyn-Thomas, H., Boyd, N. F., & Till, J. E. (1982). Attitudes toward quality of survival: The concept of "Maximal Endurable Time". *Medical Decision Making, 2*(3), 299-309.

The DCCT Research Group (1988). Reliability and validity of a diabetes quality-of-life measure for the diabetes control and complications trial (DCCT). *Diabetes Care*, *11*(9), 725-732.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273-286.

Torrance, G. W. (1986). Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics, 5*(1), 1-30.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making*, *21*(4), 329-334.

Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research, 7*(2), 118-133.

Tse, M. M. Y., Lo, A. P. K., Cheng, T. L. Y., Chan, E. K. K., Chan, A. H. Y., & Chung, H. S. W. (2010). Humor therapy: Relieving chronic pain and enhancing happiness for older adults. *Journal of Aging Research, 2010*(343574).

Tsuchiya, A. (2012). Distributional judgements in the context of economic evaluation. In A. M. Jones (Ed.), *The Elgar Companion to Health Economics* (2nd ed., pp. 406-414). Cheltenham, UK: Edward Elgar Publishing.

Tversky, A. (1996). Rational theory and constructive choice. In K. J. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behavior* (pp. 185-202). London and New York: Macmillian and St. Martins Press.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59*(4), S251-S278.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgement and choice. *Psychological Review, 95*, 371-384.

Tversky, A., Sattath, S., & Slovic, P. (2000). Contingent weighting in judgment and choice. In D. Kahneman, & A. Tversky (Eds.), *Choices, values and frames* (pp. 503-517). New York: Cambridge University Press and the Russell Sage Foundation.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review, 80*(1), 204-217.

Tversky, A., & Thaler, R. H. (1990). Anomalies. Preference reversals. *Journal of Economic Perspectives, 4*(2), 201-211.

Ubel, P. A. (1999). How stable are people’s preferences for giving priority to severely ill patients? *Social Science & Medicine, 49*(7), 895-903.

Unic, I., Stalmeier, P. F. M., Verhoef, L. C. G., & Van Daal, W. A. J. (1998). Assessment of the time-trade off values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Medical Decision Making, 18*(3), 268-277.

Van de Kuilen, G., & Wakker, P. P. (2006). Learning in the Allais paradox. *Journal of Risk and Uncertainty, 33*(3), 155-164.

Van den Berg, B., Al, M., Brouwer, W. B. F., Van Exel, N. J. A., & Koopmanschap, M. A. (2005a). Economic valuation of informal care: The conjoint measurement method applied to informal caregiving. *Social Science & Medicine, 61*(6), 1342-1355.

Van den Berg, B., Al, M., Van Exel, N. J. A., Koopmanschap, M. A., & Brouwer, W. B. F. (2008). Economic valuation of informal care: Conjoint analysis applied in a heterogeneous population of informal caregivers. *Value in Health, 11*(7), 1041-1050.

Van den Berg, B., Bleichrodt, H., & Eeckhoudt, L. (2005b). The economic value of informal care: A study of informal caregivers' and patients' willingness to pay and willingness to accept for informal care. *Health Economics, 14*(4), 363-376.

Van den Berg, B., Brouwer, W. B. F., & Koopmanschap, M. A. (2004). Economic valuation of informal care: An overview of methods and applications. *The European Journal of Health Economics, 5*(1), 36-45.

Van den Berg, B., Brouwer, W. B. F., Van Exel, N. J. A., & Koopmanschap, M. A. (2005c). Economic valuation of informal care: The contingent valuation method applied to informal caregiving. *Health Economics, 14*(2), 169-183.

Van den Berg, B., Brouwer, W. B. F., Van Exel, N. J. A., Koopmanschap, M. A., Van den Bos, G. A. M., & Rutten, F. F. H. (2006). Economic valuation of informal care: Lessons from the application of the opportunity costs and proxy good methods. *Social Science & Medicine, 62*(4), 835-845.

Van den Berg, B., & Ferrer-i-Carbonell, A. (2007). Monetary valuation of informal care: The well-being valuation method. *Health Economics, 16*(11), 1227-1244.

Van Exel, N. J. A., Brouwer, W. B. F., Van den Berg, B., & Koopmanschap, M. A. (2006). With a little help from an anchor: Discussion and evidence of anchoring effects in contingent valuation. *Journal of Socio-Economics, 35*(5), 836-853.

Veenhoven, R. (1984). The concept of happiness. In R. Veenhoven (Ed.), *Conditions of happiness*. Dordrecht, The Netherlands: Kluwer Academic.

Veenhoven, R. (1991). Questions on happiness. Classical topics, modern answers, blind spots. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective wellbeing, an interdisciplinary perspective* (pp. 7-26): Pergamon Press.

Veenhoven, R. (1993). *Happiness in nations. Subjective appreciation of life in 56 nations 1946-1992*. Rotterdam: Erasmus University Rotterdam.

Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care *Journal of Happiness Studies, 9*(3), 449-469.

Vilagut, G., Ferrer, M., Rajmil, L., Rebollo, P., Permanyer-Miralda, G., Quintana, J. M., *et al.* (2005). The Spanish version of the Short Form 36 Health Survey: a decade of experience and new developments. *Gaceta Sanitaria, 19*(2), 135-150.

Vitaliano, P. P., Young, H. M., & Zhang, J. (2004). Is caregiving a risk factor for illness? *Current Directions in Psychological Science, 13*(1), 13-16.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, New Jersey: Princeton University Press.

Walters, S., & Brazier, J. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes, 1*(4).

Ware, J. E., Kosinski, M., & Dewey, J. E. (2000). *How to score version 2 of the SF-36 Health Survey*. Lincoln, RI: QualityMetric, Incorporated.

Ware, J. E., Kosinski, M., & Keller, S. D. (1994). *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, Massachusetts: The Health Institute, New England Medical Center.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 344*(3), 220-233.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*(6), 473-483.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *The SF36 Health Survey Manual and Interpretation Guide*. Boston, Massachusetts: The Health Institute, New England Medical Center.

Watson, D., Clark, L. A., & Tellegan, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Watson, D., Pichler, F., & Wallace, C. (2010). *Second European Quality of Life Survey: Subjective well-being in Europe*. Dublin: European Foundation for the Improvement of Living and Working Conditions.

Weinstein, M. C. (1990). Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care, 6*(1), 93-103.

Weinstein, M. C., & Stason, W. B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine, 296*(13), 716-721.

Weinstein, M. C., Torrance, G. W., & McGuire, A. (2009). QALYs: The basics. *Value Health, 12 Suppl 1*, S5-9.

Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: The QALY and utilities. *British Medical Bulletin, 96*(1), 5-21.

Wilig, R. D. (1976). Consumer's surplus without apology. *American Economic Review, 66*(4), 589-597.

Wilkinson, R. A., & Chilton, G. (2013). Positive art therapy: Linking positive psychology to art therapy theory, practice, and research. *Art Therapy: Journal of the American Art Therapy Association, 30*(1), 4-11.

Wilson, E. O. (1992). *Sociobiology*. Cambridge, MA: Harvard University Press.

Willig, R. (1976). Consumer's surplus without apology. *The American Economic Review, 66*(4), 589-597.

World Health Organization (1948). Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June 1946, signed on 22 July 1946 by the representatives of 61 states (Official Records of the World Health Organization, No. 2, p. 100) and entered into force on 7 April 1948.

World Values Survey Association (2005-2008). World Values Survey. http://www.worldvaluessurvey.org.

Wurm, S., Tomasik, M. J., & Tesch-Römer, C. (2008). Serious health events and their impact on changes in subjective health and life satisfaction: The role of age and a positive view on ageing. *European Journal of Aging, 5*(2), 117-127.

Zaidi, Z. (2010). Gender differences in human brain: A review. *The Open Anatomy Journal, 2*, 37-55.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of eimultaneous equations. *Econometrica, 30*(1), 54-78.