UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

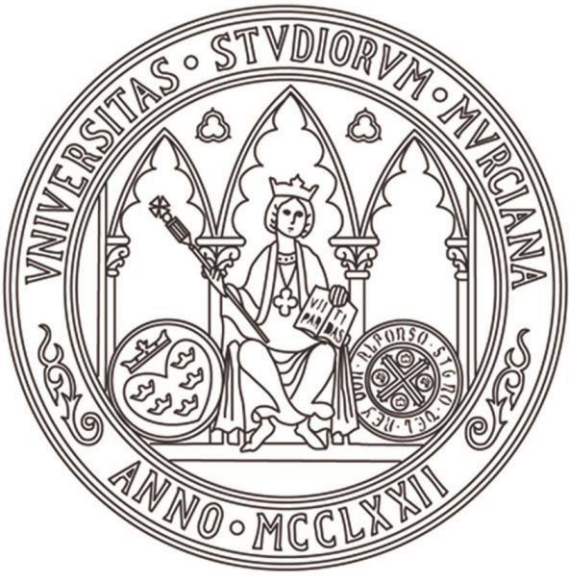
TESIS DOCTORAL

Priority Setting, Information Value and Performance Incentives in Healthcare.

Establecimiento de prioridades, Valor de la información e Incentivos al desempeño en Sanidad

D. Jorge Luis Gómez Torres

2024



UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

Priority Setting, Information Value and Performance Incentives in Healthcare.

Establecimiento de prioridades, Valor de la información e Incentivos al desempeño en Sanidad

Autor: D. Jorge Luis Gómez Torres

Director/es: D. José María Abellán Perpiñán y Jorge Eduardo Martínez Pérez

UNIVERSIDAD DE
MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD
DE LA TESIS PRESENTADA EN MODALIDAD DE COMPENDIO O ARTÍCULOS PARA OBTENER EL TÍTULO DE DOCTOR
Aprobado por la Comisión General de Doctorado el 19-10-2022

D./Dña. JORGE LUIS GÓMEZ TORRES

doctorando del Programa de Doctorado en

DOCTORADO INTERUNIVERSITARIO EN ECONOMÍA "DECIDE"

de la Escuela Internacional de Doctorado de la Universidad Murcia, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Establecimiento de prioridades.Valor de la información e Incentivos al desempeño en sanidad I

y dirigida por:

D./Dña. José María Abellán Perpiñán

D./Dña. Jorge Eduardo Martínez Pérez

D./Dña.

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Además, al haber sido autorizada como compendio de publicaciones o, tal y como prevé el artículo 29.2 del reglamento, cuenta con:

- *La autorización por escrito de los coautores de las publicaciones de que el doctorando las presente como parte de la tesis*
- *En su caso, la renuncia por escrito de los coautores no doctores de dichos trabajos a presentarlos como parte de otras tesis doctorales en la Universidad de Murcia o en cualquier otra universidad.*

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

En Murcia, a 26 de septiembre de 2024

Fdo.: Jorge Luis Gómez Torres

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe incluirse en el primer capítulo de la tesis presentada para la obtención del título de Doctor.

AGRADECIMIENTOS

Con profunda estima y reconocimiento, mi más sincera gratitud a mis directores de tesis, Dr. José María Abellán Perpiñán y el Dr. Jorge Eduardo Martínez por su enorme apoyo, compañía y sentido del humor a lo largo de todo el proceso; sin su ayuda y enseñanza, este trabajo no habría sido posible. Mi gratitud también se extiende al Dr. Fernando Ignacio Sánchez Martínez, por el apoyo que me ha brindado a lo largo de todo este período.

Finalmente, mi agradecimientos a toda mi familia, a mi mujer por su infinita paciencia y cariño, a mi suegra por su continuo apoyo en los momentos en los que he necesitado de su concurso para poder conseguir tiempo para continuar este trabajo, a mi hermana por siempre estar ahí para mí en todos los momentos duros que ha habido a lo largo del camino y a mi padre por su apoyo y consejos, que fomentaron el espíritu crítico esencial para el análisis profundo de las problemáticas en estudio.

.

”Caminante, son tus huellas, el camino y nada más;Caminante, no hay camino,se hace camino al andar.Al andar se hace el camino,y al volver la vista atrás,se ve la senda que nunca, se ha de volver a pisar.Caminante no hay caminosino estelas en la mar” *Antonio Machado*

Resumen

Esta tesis aborda tres problemáticas fundamentales desde la óptica de la economía de la salud. En el contexto actual, caracterizado por la creciente complejidad de la atención sanitaria y la salud poblacional, consideramos imperativo afrontar cuestiones específicas con un enfoque metodológico riguroso. Este acercamiento no solo pretende contribuir al acervo científico, sino también facilitar la implementación de mejoras tangibles tanto en la prestación de servicios sanitarios como en la salud general de la población. Es precisamente este principio rector el que ha guiado la concepción y desarrollo de la presente investigación doctoral.

El primero de los problemas pretende contribuir a la toma de decisiones de inversión en el Servicio Murciano de Salud. En este trabajo nos hemos centrado en desarrollar un marco de análisis de decisión multicriterio (MCDA) para la adopción de tecnologías sanitarias en un servicio regional de salud. La motivación para llevar a cabo este estudio surge de la creciente complejidad que enfrentan los responsables de tomar decisiones en el ámbito de la sanidad, donde deben equilibrar una amplia gama de factores, que incluyen los beneficios clínicos, los costos económicos y las implicaciones organizativas y sociales. En un contexto de recursos limitados y una presión cada vez mayor por parte de la demanda de servicios de salud, se requiere un enfoque más estructurado y transparente para priorizar la incorporación de nuevas tecnologías sanitarias, asegurando que las decisiones se tomen de manera fundamentada y en línea con los objetivos del sistema sanitario.

La creciente demanda de tecnologías innovadoras, junto con el envejecimiento de la población y el aumento de la cronicidad de las enfermedades, genera una presión adicional sobre los presupuestos sanitarios. Esto ha llevado a los gestores de la salud a buscar métodos que permitan evaluar las tecnologías de manera más efectiva, no solo considerando los beneficios clínicos, sino también otros criterios relevantes, como el impacto económico y la viabilidad de implementación. Hasta la fecha, los métodos tradicionales de evaluación se han centrado principalmente en los resultados clínicos y los costos, pero reconocemos la necesidad de adoptar enfoques más amplios y holísticos que incorporen una mayor diversidad de criterios. Esta

necesidad nos ha impulsado a explorar el potencial del MCDA como una herramienta para estructurar y facilitar estas decisiones complejas.

Nuestro enfoque se basa en la adaptación de marcos previos de MCDA, como el modelo EVIDEM y otros utilizados internacionalmente, pero ajustado a las particularidades del sistema sanitario español. En este sentido, hemos realizado una revisión exhaustiva de la literatura y una validación a través de consultas con expertos y profesionales del ámbito sanitario para seleccionar los criterios más adecuados. El objetivo final es proporcionar a los tomadores de decisiones una herramienta útil que permita valorar las tecnologías sanitarias desde una perspectiva multidimensional, facilitando la priorización de aquellas que ofrecen un mayor valor añadido para el sistema en su conjunto, considerando tanto aspectos clínicos como económicos y organizativos.

Para la metodología de nuestro estudio, hemos seguido un enfoque sistemático basado en el desarrollo y validación de un marco de MCDA. En primer lugar, realizamos una revisión exhaustiva de los modelos de MCDA existentes en el ámbito sanitario, con especial atención a aquellos aplicados a la evaluación de tecnologías sanitarias, como el modelo EVIDEM y el marco del Centro Belga de Conocimientos para la Atención Sanitaria (KCE). A partir de esta revisión, seleccionamos una lista preliminar de criterios que consideramos relevantes para la evaluación de tecnologías. Posteriormente, organizamos un grupo de discusión compuesto por profesionales de la salud y gestores sanitarios, con el fin de validar y ajustar estos criterios a las necesidades específicas de un servicio regional de salud en España.

Una vez establecidos los criterios definitivos, que incluyeron dominios como la necesidad de intervención, los resultados clínicos esperados, el conocimiento disponible sobre la tecnología, el impacto económico y la viabilidad de implementación, llevamos a cabo una encuesta para asignar pesos a estos criterios. Utilizamos dos muestras: una de profesionales sanitarios y otra del público general, con el fin de capturar diferentes perspectivas. Las encuestas fueron diseñadas para obtener una valoración explícita de la importancia relativa de cada criterio en la toma de decisiones, utilizando escalas de ponderación. Además, realizamos un análisis de la

consistencia de las respuestas y aplicamos técnicas estadísticas para verificar la validez interna de los resultados.

En cuanto a los resultados, observamos que tanto los profesionales sanitarios como el público general otorgaron mayor peso a los criterios relacionados con la necesidad de intervención y los resultados clínicos. Sin embargo, los profesionales sanitarios mostraron una mayor preocupación por el impacto económico de las tecnologías, mientras que el público general tendió a priorizar aspectos relacionados con la seguridad y la eficacia clínica. El dominio económico, aunque valorado por ambos grupos, fue más relevante para los gestores y profesionales. Estos hallazgos sugieren que, si bien existe un consenso en torno a la importancia de los resultados clínicos, la evaluación económica desempeña un papel crucial para los profesionales encargados de la gestión de los recursos sanitarios, destacando la utilidad del MCDA para reflejar las distintas perspectivas de los actores involucrados en el proceso de toma de decisiones.

En este sentido, uno de los aspectos más interesantes es cómo los profesionales sanitarios otorgan una mayor importancia al criterio económico, lo que subraya la necesidad de equilibrar los beneficios clínicos con las restricciones presupuestarias. Esto es especialmente relevante en un contexto de creciente presión sobre los recursos del sistema sanitario, donde es necesario priorizar aquellas tecnologías que ofrezcan un mayor valor por dinero invertido. Los resultados de nuestro estudio sugieren que, aunque la eficacia clínica sigue siendo primordial, el impacto económico y la viabilidad organizativa no deben pasarse por alto, especialmente desde la perspectiva de los tomadores de decisiones y los gestores sanitarios.

Las diferencias entre el público general y los profesionales también abren una discusión importante sobre la transparencia y la educación en la toma de decisiones sanitarias. Es probable que la población general tenga menos conciencia sobre los desafíos financieros y logísticos que enfrenta el sistema de salud, lo que refuerza la necesidad de mejorar la comunicación entre los gestores sanitarios y los ciudadanos. La implementación de marcos MCDA podría facilitar una toma de decisiones más transparente y participativa, al incorporar y equilibrar las diversas preocupaciones y prioridades de los distintos actores.

Desde nuestro punto de vista el estudio demuestra que el uso de MCDA ofrece un enfoque robusto y estructurado para priorizar la adopción de tecnologías sanitarias, permitiendo un análisis más completo que va más allá de los beneficios clínicos. Al combinar criterios clínicos, económicos y de viabilidad, este marco no solo optimiza la asignación de recursos, sino que también proporciona un proceso más transparente y justificado para los tomadores de decisiones. No obstante, es necesario seguir refinando este marco y validarlo en aplicaciones prácticas para asegurar que sea capaz de reflejar fielmente las prioridades y restricciones del sistema sanitario en su conjunto. Además, se recomienda fomentar la participación pública para alinear mejor las decisiones de adopción tecnológica con las expectativas y necesidades de la población.

El segundo de los problemas trata sobre valorar económicamente la información contenida en los prospectos de los medicamentos utilizando la metodología de valoración contingente (VC). La motivación de nuestro estudio surge de la necesidad de entender mejor el valor intrínseco que los pacientes y la población general atribuyen a la información proporcionada en estos prospectos, más allá de los beneficios directos para la salud. Tradicionalmente, se ha asumido que el valor de las intervenciones sanitarias depende únicamente de los resultados en términos de salud. Sin embargo, creemos que la información sobre riesgos y beneficios también genera utilidad para los pacientes, lo que justifica la necesidad de cuantificar ese valor.

La literatura previa ha mostrado que las preferencias de los individuos no solo están determinadas por los resultados finales de los tratamientos, sino también por características del proceso, como la accesibilidad y la claridad de la información. Los estudios anteriores que abordan la valoración de información médica, como el de Dealy et al. (2020), han estimado la disposición a pagar (DAP) por formatos estandarizados de prospectos en el contexto de la medicina estadounidense. Nuestro trabajo se basa en esta premisa, pero damos un paso más al comparar dos formatos distintos: un prospecto tradicional sin información cuantitativa y un folleto adicional con datos precisos sobre los beneficios y efectos adversos de un medicamento hipotético. A través de este enfoque, buscamos entender cómo la mejora en la comunicación de riesgos afecta la valoración de la información por parte de los usuarios. En nuestro estudio

justificamos el uso de la metodología de valoración contingente para este propósito. Esta técnica, ampliamente utilizada en la evaluación de bienes no comerciales (p.ej. el medioambiente), permite estimar la DAP por la información contenida en los prospectos de los medicamentos. Además, señalamos que la mejora de la comprensión del contenido de estos prospectos es un tema crucial en la toma de decisiones informadas por parte de los pacientes. Nos centramos en la importancia de la comunicación de riesgos de manera efectiva, presentando información sobre los beneficios y daños potenciales de los medicamentos de forma clara y comprensible, utilizando frecuencias naturales y ayudas visuales, tal como lo recomienda la evidencia más reciente en el campo de la comunicación de riesgos.

Para llevar a cabo nuestro estudio, hemos diseñado una encuesta basada en la metodología VC, que nos permite estimar la DAP de los individuos por recibir información adicional sobre los beneficios y efectos adversos de un medicamento. La muestra utilizada consta de 217 adultos de la región de Murcia, España, que fueron seleccionados de manera aleatoria para garantizar la representatividad de la población general. Los participantes fueron divididos en dos grupos: el primer grupo evaluó el medicamento con un prospecto tradicional, mientras que el segundo grupo evaluó el mismo medicamento pero acompañado de un folleto adicional que contenía información cuantitativa sobre los riesgos y beneficios, presentada mediante ayudas visuales y en frecuencias naturales.

Para la recogida de datos, empleamos dos formatos distintos de elicitación. En el primer grupo, los participantes indicaron su DAP por separado para el medicamento y el folleto complementario, utilizando una tarjeta de pagos combinada con una pregunta abierta. En el segundo grupo, los participantes indicaron su DAP por el paquete completo (medicamento más folleto), y posteriormente asignaron un porcentaje del total a la información adicional proporcionada en el folleto. Esta doble estrategia nos permitió contrastar los resultados entre ambos grupos y verificar la consistencia de las respuestas mediante pruebas de validez y fiabilidad.

En cuanto a los resultados, encontramos que la disposición a pagar por la información adicional proporcionada en el folleto varió entre 0,60 y 1 euro al mes. Los participantes del segundo

grupo, quienes valoraron el folleto como parte del paquete completo, mostraron una DAP significativamente mayor en comparación con aquellos del primer grupo, que lo valoraron por separado. Este hallazgo sugiere que el formato de elicitación influye en la valoración de la información. Además, los análisis econométricos revelaron que variables como la edad y el nivel de ingresos están positivamente correlacionadas con la DAP, lo que valida teóricamente los resultados. Sin embargo, la experiencia previa con problemas de salud cardíaca no tuvo un impacto significativo en las respuestas.

En la discusión de los resultados, destacamos que los individuos atribuyen un valor positivo a la información adicional sobre los beneficios y efectos adversos de los medicamentos, lo que confirma nuestra hipótesis inicial. Este valor intrínseco de la información no solo se refleja en la DAP por el folleto complementario, sino también en la importancia que los participantes otorgaron a contar con datos más claros y detallados. Además, observamos que el formato de elicitación afecta significativamente la valoración: aquellos participantes que valoraron el folleto como parte de un paquete integral mostraron una mayor DAP que aquellos que lo hicieron de manera separada. Este hallazgo sugiere la existencia de sesgos cognitivos inducidos por la forma de preguntar por la DAP, exacerbando la influencia de la imprecisión de las preferencias en las respuestas.

Otro punto de discusión relevante es la relación entre la edad y la DAP, que sigue un patrón en forma de U invertida. Los participantes de mediana edad parecen estar más dispuestos a pagar por la información adicional que los más jóvenes o los mayores, posiblemente debido a un mayor interés en la prevención de problemas de salud o una mayor capacidad económica. Este hallazgo, junto con la significativa correlación positiva entre el nivel de ingresos y la DAP, refuerza la validez teórica del instrumento utilizado. No obstante, la falta de impacto de la experiencia previa con problemas cardíacos o la habilidad matemática de los encuestados en las respuestas es un área que requiere una mayor exploración en estudios futuros, ya que podría reflejar una falta de conexión entre la experiencia personal y la valoración de la información proporcionada.

Desde nuestro punto de vista, nuestro estudio aporta evidencia sobre el valor intrínseco que los individuos otorgan a la información contenida en los prospectos de los medicamentos, especialmente cuando esta es presentada de forma clara y cuantitativa. Los resultados sugieren que una mejora en la comunicación de riesgos y beneficios, utilizando ayudas visuales y frecuencias naturales, no solo facilita la comprensión, sino que también incrementa la valoración económica de dicha información. Aunque este estudio ofrece importantes hallazgos, también reconocemos sus limitaciones, como el tamaño de la muestra y la necesidad de aplicar la metodología en contextos más amplios y con pacientes reales. Futuras investigaciones podrían incorporar metodologías alternativas, como experimentos de elección discreta, para profundizar en cómo los individuos valoran los atributos del proceso en la toma de decisiones sanitarias.

El tercero de los trabajos consiste en una revisión sistemática de la literatura sobre los esquemas de pago por desempeño (P4P, por sus siglas en inglés) en la atención primaria de salud, con el objetivo de identificar los elementos clave que un esquema de incentivos debe tener para mejorar de manera efectiva el rendimiento de los profesionales de la salud en el Sistema Nacional de Salud español. La motivación para llevar a cabo este trabajo surge del creciente interés en los esquemas P4P como herramientas para alinear los incentivos económicos con la mejora de la calidad asistencial. A medida que los sistemas sanitarios enfrentan mayores demandas de calidad y eficiencia, se vuelve crucial explorar mecanismos que permitan incentivar adecuadamente a los profesionales para alcanzar los objetivos establecidos por las autoridades sanitarias.

La introducción de estos sistemas en diferentes países ha generado una base significativa de evidencia sobre sus beneficios y limitaciones. En particular, el Reino Unido, con su programa *Quality and Outcomes Framework* (QOF), ha sido pionero en la implementación de P4P a gran escala en atención primaria. Sin embargo, la evidencia sobre el impacto de estos sistemas en los resultados de salud y en la equidad sigue siendo variada. En España, a pesar de que no existe un programa nacional formal de P4P en atención primaria, la creciente presión sobre el gasto sanitario y la necesidad de mejorar los resultados de salud de la población han llevado a

explorar la posible adaptación de estos esquemas en el contexto español. Nuestro estudio, por tanto, se enfoca en revisar las experiencias internacionales para identificar lecciones valiosas que puedan guiar el diseño de esquemas de P4P aplicables al sistema sanitario español.

En la primera parte del trabajo, presentamos un marco conceptual para comprender las relaciones esperadas entre los incentivos económicos y la calidad asistencial. Posteriormente, formulamos una pregunta de investigación clara utilizando la técnica PICO para guiar nuestra revisión sistemática. La búsqueda bibliográfica se realizó en bases de datos como MEDLINE, Scopus y PubMed, seleccionando estudios empíricos publicados a partir de 1999 que analizan el impacto de los esquemas P4P en la atención primaria. A lo largo del proceso de selección, nos hemos basado en las directrices PRISMA para garantizar la calidad y la transparencia de nuestra revisión. Esto nos permitió identificar los estudios más relevantes que proporcionan evidencia sobre los esquemas P4P en contextos sanitarios comparables al nuestro.

En cuanto a la metodología, hemos llevado a cabo una revisión sistemática siguiendo las directrices PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) para garantizar la exhaustividad y transparencia del proceso. Inicialmente, realizamos una búsqueda exhaustiva en las bases de datos MEDLINE, Scopus y PubMed, utilizando términos clave como "pago por desempeño" (P4P), "atención primaria" e "incentivos". Limitamos nuestra búsqueda a estudios publicados desde 1999, tanto en inglés como en español, y centrados en sistemas de salud públicos, aunque también consideramos algunos estudios relevantes en contextos privados. Aplicamos criterios de inclusión rigurosos: seleccionamos únicamente estudios que presentaban evidencia empírica sobre los resultados del desempeño de los profesionales de atención primaria bajo esquemas P4P.

Durante el proceso de selección, eliminamos duplicados y revisamos los resúmenes y textos completos de los estudios relevantes. En una primera ronda, seleccionamos 177 estudios, que luego fueron sometidos a una segunda revisión más detallada por parte de tres investigadores. Solo aquellos artículos que obtuvieron consenso entre al menos dos revisores fueron incluidos en la revisión final, lo que resultó en un total de 120 estudios que fueron analizados en profundidad. Para organizar y analizar los resultados, clasificamos los estudios en cinco

categorías clave basadas en el informe del *Quality and Outcomes Framework* (QOF) del Reino Unido: impacto en los pacientes, organización, consecuencias no deseadas, retirada de indicadores, y coste-efectividad.

Los resultados de nuestra revisión revelan varios hallazgos importantes. En primer lugar, muchos estudios no encuentran una relación causal directa entre los pagos por desempeño y la mejora en los resultados de salud de los pacientes. Sin embargo, en algunos casos específicos, los esquemas P4P han mostrado mejoras en indicadores de procesos clínicos, como la gestión de enfermedades crónicas. En términos de equidad, se observó que los incentivos a veces generan efectos no deseados, como la exclusión de pacientes más complejos para alcanzar los objetivos del esquema. Además, encontramos que los incentivos que combinan elementos individuales y colectivos, y que se basan en resultados absolutos, tienden a ser más efectivos. Sin embargo, los estudios también señalan que la eficacia de los esquemas P4P depende en gran medida del contexto organizativo y la estructura de incentivos utilizada, lo que subraya la importancia de diseñar estos esquemas de forma adaptada a las particularidades de cada sistema de salud.

En la discusión de nuestros hallazgos, es evidente que los esquemas P4P tienen el potencial de mejorar ciertos aspectos de la atención primaria, pero su efectividad depende en gran medida del diseño y la implementación de los incentivos. Uno de los principales puntos de debate es la falta de una relación causal consistente entre los incentivos y los resultados de salud. Si bien algunos estudios muestran mejoras en indicadores de procesos clínicos, como el control de enfermedades crónicas, otros no encuentran un impacto significativo en los resultados de salud a largo plazo. Esto sugiere que los incentivos económicos por sí solos no son suficientes para generar cambios sostenibles y que deben estar acompañados de otros mecanismos, como el apoyo organizativo, la formación continua y la retroalimentación.

Una de las implicaciones más relevantes para la implementación de P4P en el sistema de salud español es la necesidad de adaptar estos esquemas al contexto local, teniendo en cuenta la diversidad de los profesionales de la atención primaria y las características del sistema sanitario. Nuestros resultados subrayan la importancia de diseñar incentivos que no solo se basen en indicadores de resultados absolutos, sino que también consideren el riesgo y la complejidad de

los pacientes atendidos. Además, es crucial que los incentivos sean percibidos como justos y adecuados por los profesionales, ya que, de lo contrario, podrían generar efectos no deseados, como la exclusión de pacientes más complejos o el "gaming" de los indicadores para maximizar las recompensas.

CONTENTS

1.INTRODUCTION **1**

General setting of the thesis 1

Overview of the chapters 9

2.DESIGN OF A MULTIPLE CRITERIA DECISION ANALYSIS FRAMEWORK FOR PRIORITIZING HIGH-IMPACT HEALTH TECHNOLOGIES IN A REGIONAL HEALTH SERVICE **19**

Introduction 21

The multiple criteria decision analysis framework 23

Methodology 25

Selection and Structuring of the criteria 25

Weighting of the criteria 26

Results 28

Selection of the criteria 28

Weighting of the criteria 29

Discussion 34

Conclusions 37

Appendix 1 39

Appendix 2. Glossary of domains, criteria and subcriteria 40

Appendix 3. Some examples of the weighting procedure 43

3.THE INTRINSIC VALUE OF THE INFORMATION CONTAINED IN MEDICINE LEAFLETS **44**

Introduction 45

Methods 47

General outline of the survey design 47

The sample 48

The questionnaire 48

Willingness to pay questions: elicitation procedure 50

Hypothesis testing and internal consistency 52

Statistical methods 54

Results 57

Characteristics of the sample 57

Willingness to pay estimates 57

Feasibility, test-retest reliability and validity 63

Discussion 64

Appendix 4 70

4.PAY-FOR-PERFORMANCE IN PRIMARY HEALTHCARE: A REVIEW OF ITS STRENGTHS AND WEAKNESSES 74

Introduction 75

Methods 78

Results 81

Impact on patients 91

Organization 92

Unintended consequences 93

Withdrawal impact of indicators 93

Cost-effectiveness 93

Discussion 94

Conclusions 97

First response: a framework for prioritizing health technologies 97

Second response: an estimation of the monetary value of the information contained in medicine leaflets 100

Third response: pay-for-performance in Primary Care 103

A final corollary 105

Appendix 5 106

5.REFERENCES 109

TABLES AND FIGURES

Table 1. Criteria of the MCDA resulting from the focus group 29
Table 2. Weights of the domains, criteria and subcriteria from the two subsamples 30
Table 3. Weights (%) of domains, criteria and subcriteria for the MCDA. 33
Table 4. Brief description of the regressors used. 55
Table 5. Characteristics of the sample. 57
Table 6. Willingness to pay for the complementary brochure (€). 58
Table 7. Results of the regressions of the WTP for the brochure, WTP(i). 59
Table 8. Willingness to pay for the medicine (€). 60
Table 9. Results of the OLS regressions of the WTP for the medicine, WTP(m). 62
Table 10. Test-retest reliability (correlations). 63
Table 11.Search strategy 78
Table 12. Summary of the selected articles. 83

Figure 1. Histograms of the domains' weights from each subsample 31
Figure 2. Weights assigned to the domains by each subsample 32
Figure 3. Payment card for the valuation of the medicine (*) 51
Figure 4. Payment card for the valuation of the brochure (Group 1) 52
Figure 5. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources 80

Table A 1. Composition of the discussion group responsible for selecting domains and criteria 39
Table A 2. Sociodemographic characteristics of the general population sample 39

Figure A 1. Example of distribution of the 100 points among the 5 domains. 43
Figure A 2. Weighting task of the criteria of the "Results of the intervention" domain. 43
Figure A 3. Card 1: What is a heart attack? 70
Figure A 4. Card 2:What is a digestive bleed? 71
Figure A 5. Histograms of the Willingness to pay (WTP) for the information (i.e. the 'brochure') in each of the groups. 72
Figure A 6. Histogram of the VAS responses in Group 2: percentages of the WTP for the drug that is attributable to the information contained in the 'brochure'. 72
Figure A 7. The "traditional" leaflet. 73
Figure A 8. The complementary brochure. (Additional information on side effects) 73
Figure A 9. Conceptual Framework 106
Figure A 10.PICO Analysis (Population, Intervention, Comparison, Outcome) 108

1.INTRODUCTION

This thesis focuses on three different, though interconnected, issues from the perspective of health economics. In an increasingly complex reality such as healthcare and public health, it is essential, in our view, to tackle specific problems with a rigorous approach that not only contributes to scientific knowledge but also facilitates the implementation of improvements in healthcare provision and tangible benefits for public health. It is precisely this general principle that has guided the development of this thesis. As a preliminary step before delving into the three problems that will be the subject of study in this thesis, we will review, in the coming pages, the perspective (health economics), the context (the welfare state), and the institutional framework (the National Health System and the Murcian Health System) in which we will make our proposals.

General setting of the thesis

The etymological origin of the word "economy" traces back to ancient Greek. The term derives from two Greek words: "οἶκος" (oikos), meaning "house" or "home," and "νόμος" (nomos), meaning "custom," "law," or "administration." Together, these words form "οἰκονομία" (oikonomia), which literally translates to "household management" or "home administration." This etymology reflects the original concept of economy in ancient Greece, primarily referring to the efficient management of household resources. Xenophon, in his work "Oeconomicus" (4th century B.C.), was one of the first to use this term in a broader context, discussing the administration of property and resources (Leshem, 2013). Over time, the meaning of "economy" expanded beyond the domestic sphere to encompass resource management at the city-state level and eventually on national and international scales, evolving into the discipline we know today.

One of the fundamental principles of economics is the recognition that while human needs are infinite, the resources available to satisfy these needs are finite. This discrepancy between unlimited desires and finite resources poses an inherent challenge: how to effectively allocate

scarce resources across a variety of possible uses to maximize collective welfare. Economics, as a science, addresses precisely this allocation problem and seeks the most efficient methods for managing and utilizing available resources. Resource allocation in any economy involves making decisions about production, consumption, and distribution. In this context, economics provides analytical tools and models to evaluate available choices and their potential impacts. Without economics and its analytical approach, it would be difficult to systematically address issues of scarcity and needs, let alone design effective interventions that promote sustainable and equitable development. This discipline provides the necessary framework to balance infinite desires with the realities of limited resources, always aiming to improve the quality of life for the population.

The welfare state is a central concept in economic and social policy, referring to a system in which the government plays a key role in protecting and promoting the economic and social well-being of its citizens. This system is based on the premise that the state must ensure certain minimum living standards and access to essential services for all its citizens, including health, education, and economic security. The origins of the modern welfare state can be traced back to the late 19th and early 20th centuries. One of the earliest examples was the introduction of social insurance in Bismarck's Germany in the 1880s, which included health, accident, and pension insurances for workers (Kuhnle and Sander, 2010). This "Bismarckian" model gradually spread to other European countries. However, the welfare state as we know it today primarily developed after World War II. The Great Depression of the 1930s and the aftermath of the war created a political consensus in many Western countries about the need for greater state intervention to ensure social and economic stability. The 1942 Beveridge Report in the United Kingdom, which proposed a comprehensive social security system, was particularly influential in this regard (Briggs, 1961).

During the 1950s and 1960s, known as the "golden age" of the welfare state, many Western countries significantly expanded their social protection systems. This period saw the introduction or expansion of universal health care systems, free public education, public pensions, and other forms of social security. Esping-Andersen (1990) identified three main

models of the welfare state that emerged during this period: the social democratic model (typical of Nordic countries), the conservative-corporatist model (common in Continental Europe), and the liberal model (characteristic of Anglo-Saxon countries).

Starting in the 1970s, the welfare state began to face significant challenges. The oil crisis of 1973, rising unemployment, and slowing economic growth put pressure on public budgets. This, combined with demographic changes such as the aging population, led to debates about the sustainability of the welfare state. During the 1980s and 1990s, many countries implemented reforms to control costs and improve the efficiency of their welfare systems (Pierson, 2001).

In recent decades, the welfare state has continued to evolve in response to new challenges. Globalization, shifts in the labor market, digitalization, and new social risks have prompted a rethinking and adaptation of welfare policies. Some scholars have discussed the concept of a "Social Investment State" that emphasizes investing in human capital and preventing social risks, rather than merely providing compensation after problems occur (Hemerijck, 2013). The financial crisis of 2008 and, more recently, the COVID-19 pandemic have reignited debates about the role of the state in the social and economic protection of its citizens. These events have highlighted both the importance of robust social protection systems and the challenges of maintaining them in a context of increasing inequality and economic and social change (Greve, 2021).

Health is a fundamental component of the welfare state, reflecting the premise that access to quality medical services is a universal right, not a privilege reserved for those who can afford it. This principle aims not only to improve the quality of life of individuals but is also considered essential for the stability and productivity of society as a whole. By ensuring universal access to healthcare, welfare states help prevent the spread of diseases, reduce premature mortality, and enable all citizens, regardless of their economic status, to effectively contribute to the economy. The inclusion of health in the welfare state traces back to the late 19th and early 20th centuries. In Germany, Otto von Bismarck introduced the first mandatory health insurance system in 1883, often considered the precursor to modern public health systems (Kuhnle & Sander, 2010). This

model gradually spread to other European countries, laying the groundwork for health as a fundamental social right.

The rationale for including health in the welfare state is based on several principles. First, there is an ethical argument that health is a fundamental human right and that access to healthcare should not depend on an individual's ability to pay. This principle was enshrined in the Universal Declaration of Human Rights in 1948. Second, there is an economic argument: a healthy population is more productive and contributes more to economic growth. Additionally, public provision of healthcare can correct market failures in the health sector, such as information asymmetry and externalities (Arrow, 1963).

The evolution of the role of health in the welfare state has been significant throughout the 20th century. After World War II, many European countries established universal or near-universal healthcare systems. The National Health Service (NHS) of the United Kingdom, founded in 1948, became a influential model of a tax-funded, universally accessible healthcare system (Klein, 2013). In the United States, although a universal system was not established, programs like Medicare and Medicaid, introduced in 1965, significantly expanded the government's role in healthcare provision. From the 1970s onward, the welfare state, including healthcare systems, began facing significant challenges. The aging population, rising healthcare costs due to technological advancements, and economic crises put pressure on public budgets. This led to debates about the sustainability of public healthcare systems and various reforms in many countries (Pierson, 2001). In recent decades, there has been a growing recognition of the importance of the social determinants of health. The Black Report in the United Kingdom (1982) and subsequently the work of Marmot have highlighted how social and economic inequalities translate into health disparities (Marmot, 2005). This has led to a broader approach in health policy that extends beyond healthcare provision to address factors such as education, employment, and living conditions.

Looking to the future, the role of health in the welfare state continues to evolve. Challenges include managing chronic diseases, adapting to new medical technologies, promoting health and disease prevention, and seeking ways to improve efficiency without compromising equity.

Additionally, the COVID-19 pandemic has underscored the importance of robust public health systems and has reignited debates about the state's role in protecting public health (McKee and Stuckler, 2020).

Health economics, as a formal discipline, emerges at the intersection of economics and public health, focusing on optimizing resources and maximizing efficiency in the delivery of health services. Its development as an academic field solidified in the 1960s, marked by pioneering publications that applied economic theories to health contexts. A fundamental milestone in this development is the work of Arrow (1963), which discusses the peculiarities of the health market, highlighting how uncertainty and informational asymmetry justify state intervention in healthcare markets.

In the subsequent decades, health economics rapidly developed as an interdisciplinary field that combines elements of economics, public health, epidemiology, and health policy. Michael Grossman significantly contributed to the field with his health demand model, published in 1972, which conceptualized health as a capital good in which individuals can invest (Grossman, 1972).

During the 70s and 80s, the field expanded to address topics such as the economic evaluation of health interventions, the analysis of health systems, and the study of socioeconomic determinants of health. Alan Williams' work on Quality-Adjusted Life Years (QALYs) in 1985 laid the groundwork for developing economic evaluation methods in health (Williams, 1985).

In the 1990s, health economics gained prominence in public policy formulation. The 1993 World Bank report marked a milestone by applying economic principles to global health planning (World Bank, 1993). Simultaneously, researchers from the RAND Health Insurance Experiment provided empirical evidence on the impact of different insurance structures on the utilization of health services and health outcomes (Newhouse, 1993).

In the 21st century, health economics has broadened its scope with the incorporation of new methodologies and perspectives, such as applied behavioral economics (Thaler et al, 2008; Loewenstein et al, 2007; Volpp et al, 2011) to health and the analysis of big data in health decision-making (Dash et al, 2019; Jee and Kim, 2013). The relevance of health economics has

intensified in recent decades, given the rising health costs and the need for more equitable and efficient health systems. Through methods such as economic evaluation, microsimulation (Basu and Meltzer, 2005), and decision analysis (Weinstein et al, 2003), health economics provides essential tools to inform public policy and optimize health outcomes.

The origins of economic evaluation in health date back to the 1960s, coinciding with the development of health economics as a discipline. However, it was in the 1970s that it began to gain prominence, driven by the growing concern about rising healthcare costs and the need to justify investments in health. One of the pioneering works in this field was by Klarman et al. (1968), who conducted a cost-benefit analysis of hemodialysis programs. In turn, Fanshel and Bush (1970) introduced the idea of using a health status index as a measure of healthcare output. This index is regarded as a precursor to the concept of Quality-Adjusted Life Year (QALY), which was later coined by Zeckhauser and Shepard (1976). During the 1980s, economic evaluation in health experienced rapid methodological development, and QALYs would become a fundamental measure in health economic evaluation. This period also saw the publication of significant methodological works, such as Drummond et al. (1987), which established frameworks for the systematic conduct of economic evaluations in health.

Economic evaluation in health includes several types of analysis, each with its own characteristics and applications:

1. Cost-Benefit Analysis (CBA): This type of analysis values both costs and benefits in monetary terms. Although theoretically the most comprehensive, its application in health is limited due to the ethical and practical difficulties of assigning monetary values to health outcomes.

2. Cost-Effectiveness Analysis (CEA): Compares the costs of interventions with their outcomes measured in natural units (e.g., life years gained, cases prevented). It is the most commonly used type of analysis in health economic evaluation.

3. Cost-Utility Analysis (CUA): A specialized form of CEA that uses QALYs as the outcome measure. This approach allows for comparison of interventions with different types of health benefits.

In recent decades, economic evaluation in health has gained increasing importance in health policy decision-making. Many countries have established specialized agencies, such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom, that use economic evaluation as part of their decision-making processes regarding the incorporation of new health technologies (Drummond et al., 2015).

The recent evolution of economic evaluation in health has seen the incorporation of new methods and approaches. For example, the use of microsimulation models and model-based decision analysis has allowed addressing the complexity of many health interventions (Briggs et al., 2006). Additionally, there has been a growing recognition of the importance of considering equity in economic evaluations, leading to the development of methods to incorporate equity considerations into analyses (Cookson et al., 2017). In addition, other approaches complementary to economic evaluation have been included in the toolbox of methodologies that analysts can use to help decision-makers set priorities in healthcare. One such approach is Multi-Criteria Decision Analysis (MCDA), a set of methods that helps identify priority actions by weighing the relative importance of each assigned criterion (Devlin and Sussex, 2011).

The welfare state in Spain has a complex and relatively late history compared to other European countries. Its origins can be traced back to the early 20th century, but its full development did not occur until the end of the century, following the democratic transition. The first steps towards a welfare state in Spain were taken at the beginning of the 20th century with the creation of the National Institute of Social Welfare in 1908, which introduced the first social insurance programs (Comín, 1996). However, for much of the 20th century, the development of the Spanish welfare state was limited and fragmented, particularly during the Franco regime (1939-1975). The rationale for establishing and expanding the welfare state in Spain, as in other countries, was based on the need to provide social protection, reduce inequalities, and ensure

certain basic social rights. Additionally, in the Spanish context, the development of the welfare state was seen as a way to modernize the country and align it with European standards after decades of isolation during Francoism (Moreno, 2001).

The real expansion of the Spanish welfare state began with the democratic transition following Franco's death in 1975. The 1978 Constitution established the framework for a social and democratic state under the rule of law, guaranteeing social rights such as education, health, and social security. During the 1980s, there was a rapid expansion of social services, public education, and the national health system (Guillén and León, 2011).

The National Health System (SNS) in Spain represents a crucial milestone in the country's health policy history, marking a commitment to universal and free health for all Spanish citizens. Its formal establishment occurred in 1986 with the approval of the General Health Law, although its development began much earlier and continues to evolve. The 1986 General Health Law was a decisive step towards universalizing access to health. This law integrated all public health services into a single system, primarily funded through taxes, and established the decentralized management of the health system, giving an important role to the autonomous communities. This organization allowed health services to be adapted to local needs and improved the efficiency and accessibility of the system.

The SNS is based on principles of universality, free access, and equity. All residents in Spain have the right to free healthcare, funded by the state. This model ensures that there are no differences in access to health services based on income, employment, or personal economic situation. Additionally, the decentralized management has enabled a more effective and tailored implementation of health services, though it has also generated challenges in terms of coordination and equity among the different autonomous communities. The decentralization process was formalized mainly through the General Health Law of 1986, which established a framework for transferring health competencies from the central government to the autonomous communities. This process was carried out progressively, starting with communities that already had statutes of autonomy and a previously developed health system, such as Catalonia and the Basque Country, and continuing with the other communities until the autonomous map was

complete. The autonomous communities have the responsibility to organize and manage health resources, which includes planning health infrastructure, managing human resources, and administering hospitals and health centers. However, decentralization has also presented significant challenges. One of the main challenges has been variability in the quality and type of services offered among different autonomous communities (García-Altés et al, 2008; Cantarero, 2008; Urbanos, 2016). This variability has led to what some critics call "multi-speed healthcare," where the quality of service can significantly depend on the region in which one resides. Additionally, the financing of health services has been a source of tension, as the communities largely depend on funds allocated by the central government, which may not always adequately adjust to regional needs or costs.

The Murcian Health System (SMS) is the entity responsible for managing and administering public health services in the Region of Murcia, Spain. This system is part of the Spanish National Health System and adheres to the principles of universality, free access, and equity that characterize the Spanish public health system. Health management in Murcia, as in other autonomous communities, has been deeply influenced by the decentralization process initiated in the 1980s, which granted health competencies to the autonomous communities.

The origins of the Murcian Health System date back to the Law 4/1994, of July 26, on Health of the Region of Murcia, which established the foundations for creating a distinct health system within the autonomous community. This system was organized to provide comprehensive coverage to all citizens of the Region of Murcia, ensuring access to quality health services. With the transfer of health competencies from the State to the Autonomous Community of Murcia in 2001, the creation of the SMS was formalized, allowing the region to autonomously manage its health care.

Overview of the chapters

The structure of this thesis, organized around three core chapters, reflects a comprehensive and multidimensional approach to addressing key issues in health economics. The common thread

of the three chapters is how to improve efficiency and quality of care in the health system. Each chapter addresses a different issue attached to this aim, applying a different methodology.

The first article, which develops a multi-criteria decision analysis (MCDA) framework for prioritizing health technologies in the Murcian Health System, establishes the methodological and conceptual foundation for evaluating complex decisions in healthcare management. This framework integrates various evaluation criteria, such as clinical effectiveness, costs, and ethical considerations, facilitating more informed and equitable decision-making in contexts of limited resources. This perspective directly aligns with the goal of improving efficiency and equity in resource allocation within health systems, a critical aspect of the sustainability of any healthcare system.

The second article, focused on the intrinsic value of the information contained in drug leaflets, complements the analysis of the first by introducing a key dimension of information access in healthcare decision-making. By employing the contingent valuation methodology, it explores the perceptions and preferences of patients and consumers, contributing to the understanding of how pharmacological information influences both individual and collective health decisions. This approach links to the importance of transparency and patient empowerment in the context of health technology prioritization, as informed decision-making is fundamental at both the institutional and individual levels.

The third article, a systematic review of pay-for-performance (P4P) incentive schemes in primary care, provides a bridge between the first two approaches by focusing on incentive mechanisms to improve performance and the quality of healthcare services. By identifying international best practices and proposing some features that a hypothetical incentive scheme for the Primary Care system in Spain should have, this article deepens the analysis of how incentive policies can align with the efficient prioritization of resources, improve quality of care, and promote informed decision-making. This analysis brings the discussion full circle by connecting strategic decisions in the management of health resources and technologies with the implementation of incentives that foster better health outcomes.

Overall, the interconnection between these three articles lies in their systematic and coordinated approach to improving health systems, addressing both the macroeconomic management of health resources and the microeconomic dimension of patient and healthcare professional behavior. The thesis offers a comprehensive vision that seeks not only to optimize the allocation of health technologies and resources but also to improve access to relevant information and promote incentive schemes that enhance efficiency and quality of care, with direct applicability to the Spanish context.

Once the foundations are clearly established from the perspective, scope, and institutional environment in which the problems will be addressed, we will briefly develop each one of them. The first problem we aim to address is developing a MCDA framework for prioritizing high-impact health technologies in the SMS. MCDA is a sub-discipline of operations research that explicitly deals with multiple criteria in decision-making environments. It is a methodology that assists decision-makers in evaluating and choosing among alternatives when faced with several, often conflicting, criteria (Belton and Stewart, 2002). MCDA emerged as a formal field in the 1960s, with pioneering works such as those by Roy in France and Keeney and Raiffa in the United States. Roy developed the ELECTRE method (Elimination and Choice Expressing Reality), one of the first systematic approaches to addressing multi-criteria decision problems (Roy, 1968). Meanwhile, Keeney and Raiffa laid the theoretical groundwork for multi-objective decision analysis (Keeney and Raiffa, 1976). Over the years, numerous methods have been developed within the MCDA framework. Some of the most well-known include the Analytic Hierarchy Process (AHP) developed by Saaty (Saaty, 1980), the PROMETHEE method (Preference Ranking Organization Method for Enrichment Evaluations) proposed by Brans and Vincke (1985), and the TOPSIS method (Technique for Order Preference by Similarity to Ideal Solution) introduced by Hwang and Yoon (1981).

MCDA has found applications across a broad range of fields, including environmental management, urban planning, project selection, technology assessment, and public sector decision-making. Over the past few decades, it has garnered increasing attention in the field of health technology prioritization, offering a structured approach to evaluate and compare

different health interventions or technologies considering multiple criteria simultaneously, such as clinical effectiveness, cost-effectiveness, disease burden, equity, and ethical considerations (Thokala et al., 2016).

Thus, MCDA has been proposed as a complementary tool to traditional economic evaluation methods, such as cost-effectiveness analysis. Baltussen and Niessen (2006) argued that MCDA could provide a more transparent and comprehensive framework for decision-making in health, allowing the explicit inclusion of criteria that extend beyond economic efficiency. Since then, several countries have explored the incorporation of MCDA into their health technology assessment (HTA) processes, such as Thailand, where it has been used to prioritize public health interventions (Youngkong et al., 2012). However, implementing MCDA in health technology prioritization is not without challenges. These include selecting and weighting criteria, aggregating scores from different criteria, and interpreting results in the context of budget constraints and other practical considerations. Moreover, there is an ongoing debate about how to best integrate MCDA with existing HTA processes and decision-making in health systems (Marsh et al., 2014).

The objective of this first chapter is to provide a systematic and transparent tool that allows health managers to efficiently evaluate and prioritize these technologies, taking into account multiple criteria beyond cost-effectiveness, such as clinical effectiveness, safety, economic implications, and patient preferences. Traditionally, approaches to health technology assessment have focused on cost-effectiveness analysis, whose main goal is to maximize health benefits adjusted for available resources, using measures such as Quality-Adjusted Life Years (QALYs) and Disability-Adjusted Life Years (DALYs). However, this approach has been criticized for its unidimensional focus, leading to a growing acceptance of Multi-Criteria Decision Analysis in the health sector. MCDA provides a more inclusive framework, allowing for the simultaneous evaluation of multiple criteria relevant to health decision-making, such as safety, quality of life, patient-perceived outcomes, and direct and indirect costs. Authors like Marsh et al. (2014) and Thokala et al. (2016) have underscored the utility of this approach in complex decision-making

contexts, where it is necessary to integrate diverse perspectives and values from various stakeholders, including patients, health professionals, and system managers.

Moreover, MCDA has been utilized in various countries such as Canada, Belgium, and the United Kingdom, to support decisions on funding and prioritization of emerging technologies. Recent studies, like those by Goetghebeur et al. (2012) and Baltussen et al. (2019), have highlighted how MCDA allows for the more explicit incorporation of social preferences and ethical values, making it a powerful tool to enhance the transparency and legitimacy of health policy decisions. This theoretical framework is particularly relevant in decentralized health systems like Spain's, where autonomous communities, such as Murcia, have the autonomy to manage their resources and decide which technologies to adopt based on local needs.

The methodology employed in this chapter includes a review of existing MCDA frameworks and their adaptation to the context of the Murcian Health System. For this purpose, a focus group with clinical leaders and health managers was used to select relevant criteria, followed by surveys of two samples: healthcare professionals and the general population. The surveys enabled the weighting of different criteria, thus obtaining a balanced view of the preferences of both experts and citizens.

The second issue addressed in this thesis pertains to estimating the intrinsic value of the information contained in drug leaflets. The intrinsic value of information is a fundamental concept in decision theory and information economics. It refers to the value that information holds in itself, regardless of the decisions or actions that may be taken based on it. This concept is particularly relevant in the context of risk information, where knowledge can significantly influence decision-making and the well-being of individuals and organizations.

In the domain of risk information, the intrinsic value is manifested in the ability to reduce uncertainty and improve the quality of decisions. As Hirshleifer and Riley (1979) note in their seminal work, information has value insofar as it enables agents to make more informed and potentially more beneficial decisions. In the context of risks, this can translate into better assessments of potential threats and more efficient allocation of resources for mitigation or prevention.

The expected utility theory, developed by von Neumann and Morgenstern (1944), provides a framework for quantifying the value of information in risk situations. According to this theory, individuals make decisions based on the expected utility of possible outcomes. Additional information can alter the perceived probabilities of these outcomes, which in turn can change the optimal decision. The value of information, in this context, can be measured as the difference between the expected utility with and without that information. However, it is important to note that the value of information is not always positive or uniform for all individuals or situations. As Grossman and Stiglitz (1980) argue in their analysis of the value of information in financial markets, information can have different values for different agents, depending on their ability to interpret and act upon it. Furthermore, in some cases, additional information can even decrease welfare if it leads to greater anxiety or suboptimal decisions due to cognitive biases or processing limitations.

In the specific context of health risk information, the intrinsic value of information is further complicated by ethical and psychological factors. For example, Fischhoff (1995) has explored how risk communication can affect people's perceptions and behaviors, highlighting the importance not only of the information itself but also of how it is presented and contextualized. This underscores that the value of risk information depends not only on its accuracy or completeness but also on its capacity to be understood and effectively used by the recipients.

Risk communication is a crucial aspect of public health and medicine, particularly concerning the information provided to patients about medications. Drug leaflets play a fundamental role in this process, as they are often the main source of information for patients about the risks and benefits of the medications they consume. The effectiveness of risk communication in drug leaflets has been the subject of numerous studies. According to Raynor et al. (2007), many patients struggle to understand the information presented in leaflets, which can lead to misunderstandings about the risks associated with medications. This underscores the importance of presenting information in a clear, concise, and accessible manner for the general public.

A particular challenge in communicating risks through leaflets is the presentation of statistical information. Gigerenzer et al. (2007) have argued that the use of "natural risk formats," such as

natural frequencies instead of percentages, can significantly enhance patients' understanding of risks. For example, stating "1 out of 100 people experience this side effect" is easier to comprehend than "this side effect occurs in 1% of cases." In addition to clarity in presenting information, the contextualization of risks is crucial. Schwartz et al. (2009) note that providing information about the benefits of the medication alongside the risks can help patients make more informed decisions. However, this must be done in a balanced manner to avoid biasing the patient's perception in one direction or another.

Regulation also plays a significant role in risk communication in drug leaflets. In the European Union, for instance, Directive 2001/83/EC sets specific requirements for the information that must be included in leaflets, including a section on possible adverse effects. However, as Herber et al. (2014) point out, compliance with these regulatory requirements does not always ensure that the information is understood or effectively used by patients, highlighting the need for a multidisciplinary approach that combines regulatory, psychological, and communication perspectives in the design of leaflets.

In this context, this second chapter of the thesis seeks to assess citizens' willingness to pay (WTP) for additional, quantitative, and detailed information about the benefits and side effects of a hypothetical medication, aiming to improve patients' decision-making in terms of understanding risks and benefits. This work is framed within the economic literature that studies the willingness to pay for non-commercial goods, such as health information, and the application of the contingent valuation method to measure that value. Contingent valuation has been widely used in previous studies to estimate the value of health interventions and other health-related goods. For example, studies like those by Donaldson and Shackley (1997) or Pinto et al. (1998) have employed this method to value attributes not directly related to health improvement but that affect patients' well-being, such as treatment convenience or the information provided in leaflets.

In particular, the approach of willingness to pay for information about medications has been explored by authors such as Dealy et al. (2020), who assessed the WTP to switch to a standard format of presenting information in medical leaflets. This chapter expands on this line of

research by comparing different formats of information presentation, a traditional one and another that includes quantitative data on efficacy and adverse effects, following the best practices in risk communication as proposed by Gigerenzer et al. (2007) and Yamagishi (1997), who emphasize the importance of presenting information in a comprehensible and accessible manner for patients. The methodology employed in this chapter consists of a contingent valuation study based on surveys conducted on a sample of the general population in the Region of Murcia.

The third problem addressed in this thesis is related to the payment of performance incentives in primary care. For this purpose, a systematic review of Pay-for-Performance (P4P) incentive schemes applied in primary healthcare is conducted, aiming to identify best international practices and propose a new incentive scheme for the Spanish healthcare system. This chapter intends to analyze how economic incentives influence the behavior of healthcare professionals and the outcomes of healthcare, focusing particularly on the context of primary care.

Framed within the literature on incentive systems in healthcare, the chapter addresses the evolution of P4P models, which began to be widely implemented in countries like Australia, the United States, and the United Kingdom. According to Scott et al. (2009), P4P aims to align healthcare professionals' incentives with the goals of the healthcare system, improving the quality of care through rewards linked to the achievement of specific indicators. Studies such as those by Doran et al. (2006) and Campbell et al. (2009) have shown that P4P programs can improve certain clinical processes, but the results in terms of long-term population health outcomes are mixed.

This multifaceted approach to examining incentives in healthcare provides a robust framework for understanding how different incentive models impact healthcare delivery. By analyzing empirical evidence and theoretical perspectives, this chapter contributes to a more nuanced understanding of how P4P schemes can be designed and implemented effectively within the specific context of Spain’s decentralized healthcare system. This study emphasizes the need for careful consideration of local healthcare dynamics and professional motivations to ensure that incentive schemes enhance rather than undermine the quality of primary care services.

The chapter also engages in the debate over the potential adverse effects of P4P programs. Research like that of Fleetcroft et al. (2012) has shown that poorly designed incentives can lead to "tunnel vision," where professionals focus exclusively on incentivized areas, neglecting other important aspects of care. Moreover, Saint-Lary et al. (2015) highlight that incentives can create inequities, as physicians might choose to exclude complex patients who could negatively impact their performance metrics.

Another challenge of P4P is its impact on equity in healthcare. In some cases, incentives can lead to adverse patient selection, as professionals may opt not to treat more complex or poorer prognosis patients to avoid negatively impacting their performance metrics. Saint-Lary et al. (2015) warned that this phenomenon can create inequities in care, particularly in systems with low-resource patients or those with multiple comorbidities. Therefore, it is crucial that P4P schemes incorporate risk adjustment mechanisms that account for the complexity of the patients served.

The methodology used in this chapter consists of a systematic review of the literature on P4P schemes in primary care. A comprehensive search was conducted in databases such as Medline, Scopus, and PubMed, following the PRISMA guidelines for systematic reviews. Studies published from 1999 to 2019 that analyzed empirical evidence on the impact of economic incentives on primary care outcomes were included. It was decided to end the systematic search before the onset of the COVID-19 pandemic, as between 2020 and 2022, the implementation of incentive systems was modified in many cases (e.g. the Quality and Outcomes Framework incentive programme for General Practitioners in England) to concentrate resources on responding to the pandemic health emergency. After rigorous selection, 101 studies were reviewed in depth, analyzed based on their context and outcomes, with the aim of drawing relevant conclusions for the potential implementation of a P4P scheme in the Spanish healthcare system. These studies from the systematic review were supplemented with other 7 identified from other methods.

After addressing each of these three previously mentioned problems, which are presented in the format of a scientific article, general conclusions of the work performed will be developed.

These conclusions will synthesize the insights gained from the examination of each issue, providing a comprehensive overview of the implications for policy and practice in the context of Spanish healthcare.

2.DESIGN OF A MULTIPLE CRITERIA DECISION ANALYSIS FRAMEWORK FOR PRIORITIZING HIGH-IMPACT HEALTH TECHNOLOGIES IN A REGIONAL HEALTH SERVICE[1]

Abstract

Aim: This study aims to develop a framework for establishing priorities in the regional health service of Murcia, Spain, to facilitate the creation of a comprehensive multiple criteria decision analysis (MCDA) framework. This framework will aid in decision-making processes related to the assessment, reimbursement and utilization of high-impact health technologies. **Method**: Based on the results of a review of existing frameworks for MCDA of health technologies, a set of criteria was proposed to be used in the context of evaluating high-impact health technologies. Key stakeholders within regional healthcare services, including clinical leaders and management personnel participated in a focus group (n=11) to discuss the proposed criteria and select the final ones (fifteen). To elicit the weights of the criteria, two surveys were administered, one to a small sample of healthcare professionals (n=35) and another to a larger representative sample of the general population (n=494). **Results**: The responses obtained from health professionals in the weighting procedure exhibited greater consistency compared to those provided by the general public. The criteria more highly weighted were "Need for intervention" and "Intervention outcomes". The weights finally assigned to each item in the multi-criteria framework were derived as the equal-weighted sum of the mean weights from the two samples. **Conclusions**: A multi-attribute function capable of generating a composite measure (multi-criteria) to assess the value of high-impact health interventions has been developed. Furthermore, it is recommended to pilot this procedure in a specific decision context to evaluate

[1] This chapter has been published in the *International Journal of Technology Assessment in Health Care* (doi: https://doi.org/10.1017/S0266462324000205).

the efficacy, feasibility, usefulness and reliability of the proposed tool. **Keywords**: Multiple-Criteria Decision Analysis, healthcare technologies, prioritisation, Resource allocation

Introduction

The growth of healthcare expenditure poses significant challenges to resource allocation in public health systems. Demographic (aging, morbidity, disability, and proximity to death) and non-demographic determinants of healthcare spending (biomedical technology innovation, income, and rising prices in the healthcare sector), exert considerable pressure on public budgets (Baltagi et al.,2017; Dormont et al.,2006; Howdon et al., 2018; Payne et al.,2007; Smith et al.,2009; Zweifel et al.,2005). Consequently, healthcare managers face the daunting task of making decisions with substantial opportunity costs within increasingly complex and multifaceted contexts (OECD, 2011; Schmets et al., 2016).

In the European context, a value-based approach is employed to assist in public financing and pricing decisions concerning new health technologies (Oortwijn et al., 2017). For instance, the United Kingdom primarily evaluates value by comparing the cost-utility of an intervention (measured as the Incremental Cost per Quality-Adjusted Life Year gained) with an efficiency threshold (NICE, 2013). In France and Germany, however, value is determined based on the incremental therapeutic benefits and domestic reference pricing, playing cost-effectiveness a small role in the overall approach (Toumi et al., 2015; Lauenroth et al., 2017).

Furthermore, significant advancements in biomedical innovation have added complexity to the evaluation and decision-making processes within this rapidly changing environment (Okunade et al., 2002; Oliveira et al., 2006; Willemé et al., 2014). Due to potential conflicts of interest among stakeholders, there is an increasing interest in employing methodologies that systematize the criteria for assessing health technologies. The Multiple Criteria Decision Analysis (MCDA) is particularly notable in this regard, encompassing a set of methods that assist in prioritizing actions by assigning relative importance to each criterion reflecting different dimensions of a health technology's performance. These dimensions include clinical effectiveness, safety, cost, ethical considerations, and patient preferences. (Baltusen et al., 2006; Wahlster et al., 2015; Marsh et al., 2016;Devlin et al., 2011).

The aim of this study is to design an MCDA framework to inform decisions on the incorporation of high-impact technologies in the regional health service of Murcia, Spain. By 'high impact' technologies, we mean both impact on patients' health -reducing the burden of disease they bear, and/or impact on the available budget -consequently displacing other healthcare services. This high impact can be the result of low-cost and/or low-benefit technologies, indicated for large populations, as well as the result of high-cost and/or high-benefit technologies, but aimed at small populations.

The Spanish healthcare system is a highly decentralized one, with a notable degree of autonomy in how each regional health service prioritizes funding for new healthcare technologies, especially those that do not involve pharmaceuticals. Although MCDA is currently used by some Spanish regions (e.g. Catalonia uses this methodology to assess some drugs), in the Region of Murcia -a relatively small Spanish region, with just over 3% of the national population- there is currently no formalized procedure with explicit criteria for making these decisions. This lack of a standardized process results in significant differences between health areas or hospital centers.

The specific objectives are to select the criteria that will be part of the scheme, as well as to obtain the weights of each of them based on the preferences of health professionals and the general population. The task of assigning scores to each of the criteria is outside the scope of our study, so in this respect it is similar to the approach followed by Cleemput et al. (2014) in their report for the Belgian Health Care Knowledge Centre (KCE).

The next section provides a summary of the fundamental aspects of MCDA and its applications. We then elaborate on the methodologies employed to develop an MCDA framework tailored to assess high-impact health technologies within the context of a Spanish regional health service. The findings derived from the analysis are presented next, followed by a Discussion section, which precedes the final conclusions.

The multiple criteria decision analysis framework

A classical definition of MCDA is that by Keeny and Raiffa (Keeny et al., 1993), “a methodology for appraising alternatives on individual, often conflicting criteria, and combining them into one overall appraisal”. The potential of MCDA in healthcare decision-making was recognized in the 1980s, with the growing need to incorporate multiple perspectives and criteria into health technology assessment (HTA). Since then, the use of MCDA in HTA has been actively promoted, based on its potential, but also criticized, because of doubts about its suitability (Campillo-Artero et al., 2018). Nevertheless, MCDA has been widely utilized in the healthcare sector for various decision-making purposes (Thokala et al.,2016; Glaize et al., 2019), such as new technology evaluations (Husereau et al., 2010; Danner et al., 2011), assessment of orphan drugs (Sussex et al., 2013; Gilabert-Perramon et al., 2017), risk-benefit assessments (Phillips et al., 2011), hospital purchasing (Dolan 1989; Dolan, 2008; Van Tin JA et al., 2008; Pecchia et al., 2013), and establishing priority frameworks for different types of interventions (Goetghebeur et al., 2008).

Interest in using MCDA to inform decisions on public financing of new technologies has also grown in recent decades. Consequently, various guidelines have been developed based on this methodology by HTA institutions and agencies, such as Canada (Huserau et al., 2010), the United Kingdom (Devlin et al.,2011), Belgium (Cleemput et al., 2014), and Spain (Marqués-Peláez et al., 2020).

Two main modalities of MCDA are typically distinguished: qualitative MCDA and quantitative MCDA. In qualitative MCDA, technologies are evaluated through deliberation about their performance on explicitly defined criteria. In other words, a qualitative interpretation of the "performance matrix" takes place (Baltusen et al., 2019). The goal of quantitative MCDA is to obtain a global measure of the value of each technology. An overwhelming majority of studies that have utilized MCDA in HTA are of a quantitative nature (Baltusen et al., 2019).

The quantitative MCDA framework comprises three primary phases (Devlin et al., 2011): selection of criteria, weighting of criteria, and application of the framework established in the two previous phases. The selection of criteria must adhere to the requirements set forth in the

recommendation guide of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). These requirements include completeness, non-redundancy, no-overlap and preference independence (Marsh et al., 2016).

Performance for each criterion can be measured using various scales (binary, categorical, ordinal, ratio, interval, etc.). On the other hand, weighting involves eliciting stakeholders' preferences between criteria (Campillo-Artero et al., 2018). Weights reflect the "trade-offs" between criteria and are needed to combine the scores on individual criterion into a unique measure of "total value".

There are different types of methods for scoring and weighting criteria: direct methods, hierarchical methods, discrete choice methods, and matching methods (Marsh et al., 2017). The source of preferences depends on the type of decision problem. The "stakeholders" can be members of the Regulatory Committees or the Health Technology Assessment Committees, patients, clinical leaders and other health professionals, or the general public (Campillo-Artero et al., 2018).

Once the alternatives' performance is scored and the criteria are weighted, their values must be aggregated to determine which intervention generates the highest value. Aggregation can be performed using a variety of procedures (e.g. additive or multiplicative methods, regression methods), depending on the methods used to score the criteria and assign weights (Marsh et al., 2014).

Subsequently, uncertainty analysis in the MCDA framework is conducted similarly to economic evaluation studies. Sensitivity analysis should consider all sources of uncertainty (structural, stochastic, parameter, etc.), and can be deterministic or probabilistic (Broekhuuizen et al., 2015).

Methodology

Selection and Structuring of the criteria

To select the criteria that will constitute the MCDA framework, a discussion meeting was conducted with a carefully selected group of eleven organizational members who possess decision-making authority regarding the purchase and use of these technologies (referred to as decision-makers). The group included various high-ranking officials from the regional health service, as well as health area managers and other mid-level executives (more detailed information is available in Appendix 1). The meeting took place on November 26, 2021, at the facilities of the regional health service.

Prior to the meeting, the participants were provided with a list of criteria. These criteria resulted from a two-step pre-selection process conducted by the research team. Firstly, a set of criteria were selected from the latest version of the EVIDEM framework (EVIDEM, 2017). The EVIDEM (Evidence and Value: Impact in Decision-Making) framework consists of a "core model" with thirteen quantifiable criteria, grouped into five domains, supplemented by a contextual tool of six qualitative criteria and one criterion related to the opportunity costs of the intervention. Each generic criterion can also include specific subcriteria relevant to a particular therapeutic area or type of intervention.

Fourteen criteria were chosen, comprising the thirteen criteria from the "core model" and the Opportunity Cost Considerations criterion. The reason for selecting most of the criteria from the EVIDEM framework was that these criteria are generic and universally applicable (Goetghebeur et al., 2010).

Additionally, the criteria from the KCE framework were integrated, with appropriate modifications when necessary. The KCE report (Marsh et al., 2013), includes results from a survey of the general population and health decision-makers aimed to assign weights to ten criteria grouped into three categories: therapeutic needs, social needs, and the added value of the new treatment. These criteria were based on a transparent decision framework previously developed by the KCE (Polain et al., 2010). This framework was designed to enhance accountability in the realm of public healthcare benefits reimbursement, a goal closely aligned

with the objectives of our proposal. Hence, we chose to integrate some of its criteria in our framework.

The criteria thus selected were then grouped into five domains, and are those shown in Table 1, with the exceptions and qualifications indicated at the foot of the table. The precise definition of domains, criteria and sub-criteria can be found in the glossary (Appendix 2).

The dynamics of the meeting with the decision makers was as follows: First, the objective and mechanics of the meeting were explained to the participants. The domains were then voted on, followed by a debate and discussion of the results, which, if applicable, could lead to an extension or reduction of the domains. The criteria were then voted on, following the same methodology as for the domains: voting, debate and discussion and, if necessary, extension, reduction and/or relocation of the criteria. Finally, this same process was carried out with the sub-criteria included within each criterion previously selected.

It is important to emphasize that, before each vote, participants could suggest additions or modifications to the list of domains or criteria under consideration. The objective was to reach final decisions by consensus after discussing the results following each vote. In the event that consensus was not achieved, the majority rule was applied. This structured meeting format allows for active participation from decision makers, facilitating the refinement and finalization of the framework of criteria and sub-criteria to be employed in the evaluation of high-impact health technologies.

Weighting of the criteria

To obtain the weights associated with the criteria, we conducted surveys with two distinct samples: decision-makers and healthcare professionals from the Regional Health Service, and a sample drawn from the general population of the Region of Murcia, Spain. This approach allows us to compare the judgments of healthcare professionals, who possess specialized expertise, against the presumably less informed viewpoint of the general population.

A total of sixty-seven professionals received an invitation by the Regional Health Service to complete the questionnaire. Among the recipients were area managers, hospital medical

directors, coordinators, and heads of specialized services with high technological requirements (surgery, oncology, etc.). The response rate was 52% (thirty-five respondents).

A representative sample of the population (n=500) was obtained through a two-stage stratified sampling methodology. To optimize the response rate, recruitment strategies included advance contact, reminders, and appointment scheduling. As the survey was endorsed by the Health Department, high collaboration was achieved, obtaining a response rate of 99% (494 valid questionnaires). Statistics of this sample are available in Appendix 1.

Two questionnaires were designed and interfaces were programmed for this purpose, with one questionnaire tailored for each sample. The structure of each questionnaire was similar in both surveys, except for the need to include some additional information for the general public. Wording was slightly simplified in the questionnaire administered to the general population, to ensure comprehension. Both questionnaires started with an introduction to the survey's primary objective, namely, to determine the relative importance assigned by the respondents to the different criteria within the analysis framework.

The questionnaire for professionals was administered online, with the selected individuals receiving an email invitation from the Regional Health Service. For the general population sample, computer-assisted personal interviews (CAPI) were conducted at the participants' homes.

To assign weights to the domains, criteria, and sub-criteria, we utilize the allocation of 100 points, method employed in the EVIDEM framework. This method involves distributing 100 points among the domains, 100 points among the criteria within each domain, and 100 points among the sub-criteria within each criterion. Some screenshots can be seen in Appendix 3.

The weights obtained from the two subsamples were compared by means of parametric (t-test for independent samples) and nonparametric (Mann-Whitney- Wilcoxon) tests.

Results

Selection of the criteria

The initial proposal described in the previous section was presented to the eleven members of the discussion group responsible for selecting the criteria. Before voting on the domains, one of the participants suggested adding a domain that captured the availability of resources within the healthcare system to incorporate the technology under evaluation, as well as its impact on the system's organization. This proposal was accepted by consensus, and the "Feasibility" domain was added, including two criteria (see Table 1. All domains received unanimous support from the participants, except for the "Knowledge of the intervention" domain, which recorded two opposing votes.

The criteria received unanimous endorsement from the participants, with few exceptions: "Comparative safety," "Patient-perceived outcomes," "Preventive benefit," "Therapeutic benefit," and "Non-healthcare costs" received one opposing vote; the "Expert consensus" criterion was supported by eight out of eleven participants. After a brief debate, participants agreed to relocate the domain "Type of benefit" and its corresponding criteria ("Preventive benefit" and "Therapeutic outcome") as a criterion within the "Outcome of the intervention" domain.

The subcriteria that did not receive 100 percent of the votes from the attendees were "Unmet needs in HRQoL", "Change in intermediate outcomes", and "Change in HRQoL" (one opposing vote each), "Change in convenience" (three opposing votes), and "Unmet needs in convenience" (four opposing votes).

It was understood that all criteria and subcriteria were validated by the participants in the meeting, with the clarifications provided. The final criteria are as shown in Table 1.

Table 1. Criteria of the MCDA resulting from the focus group

Domains	***Criteria***	***Subcriteria***
Need for intervention	• Disease severity	• Impact on HRQoL • Impact on life expectancy
	• Affected population	---
	• Unmet needs	• In effectiveness • In HRQoL • In safety • In convenience
Outcomes of the intervention	• Comparative effectiveness	• Change in life expectancy • Change in intermediate results • Change in prevalence
	• Comparative safety	---
	• Comparative patient-reported outcomes	• Change in HRQoL • Change in convenience
	• Type of Benefit [(1)]	• Preventive benefit • Therapeutic benefit
Knowledge about the intervention	• Quality of evidence	• Validity • Relevance
	• Expert consensus	---
Economic impact	• Direct healthcare costs	---
	• Other healthcare costs	---
	• Non-medical costs	---
	• Opportunity cost and budget impact	---
Feasibility [(2)]	• Availability of resources in the system	---
	• Organizational impact	---

Source*:* Own elaboration, based on EVIDEM 10th edition (55), the KCE framework. (35), and the results of the decision-makers discussion group.

[(1)] The criterion "Type of benefit" was initially included as a domain in the proposal submitted for debate and vote. The participants in the focus group agreed to relocate it as a criterion, within the domain "Outcomes of the intervention". [(2)] The domain "Feasibility" and its two criteria were absent in the initial proposal, but were added as a result of the focus group discussion.

Weighting of the criteria

Table 2 presents the mean weights, accompanied by their standard deviation, for all the domains, criteria, and sub-criteria, obtained from each sample. In both cases, the same three domains receive the highest weightings. "Need for intervention" occupies the top position, with a weight of 28.1 percent in the general population subsample and 23.7 percent in the healthcare professionals' sample. The domain "Intervention outcomes" is ranked second (24.6 and 23.1 percent, respectively), and the third domain is "Knowledge about the intervention", (19.0 and 19.5 percent). In the general population subsample, the fourth-ranking domain is "Feasibility" (14.5 percent). Conversely, healthcare professionals place the domain "Impact on the economy" in fourth position (18.5 percent).

Table 2. Weights of the domains, criteria and subcriteria from the two subsamples

	General population		Health-care profesionals		Difference (GP – HCP)	
	Mean	St.Dev.	Mean	St.Dev.	Mean	*P-value*
I.- Need for intervention	**28,08**	**16,28**	**23,69**	**7,60**	**4,39**	**0,114**
• DIsease severity	41,10	18,52	36,31	11,55	4,79	0,132
○ Impact on HRQoL	55,53	20,46	61,31	11,56	-5,79	0,099
○Impact on life expectancy	44,47	20,46	38,69	11,56	5,79	0,099
• Affected population	31,18	15,88	36,71	9,96	-5,53	0,043*
• Unmet needs	27,71	16,86	26,97	9,01	0,74	0,797
○ In effectiveness	28,16	15,69	31,29	9,51	-3,12	0,246
○In CVRS	28,73	14,66	25,40	6,46	3,33	0,183
○In safety	22,76	12,06	24,00	6,04	-1,24	0,546
○ In convenience	20,35	13,30	19,31	5,21	1,03	0,648
II.- Outcomes of the intervention	**24,56**	**15,89**	**23,14**	**6,31**	**1,42**	**0,601**
• Comparative effectiveness	26,59	14,14	28,40	8,30	-1,81	0,455
○ Change in life expectancy	39,48	19,50	37,74	8,87	1,73	0,602
○ Change in intermediate results	31,10	17,24	29,54	7,39	1,55	0,597
○ Change in prevalence	29,43	17,08	32,71	10,00	-3,29	0,262
• Comarative safety	23,97	12,96	23,63	6,23	0,34	0,877
• Comparative patient reported outcomes	25,31	15,25	24,31	6,83	1,00	0,702
○ Change in HRQoL	59,18	19,59	61,97	9,88	-2,80	0,403
○ Change in convenience	40,82	19,59	38,03	9,88	2,80	0,403
• Type of benefit	24,13	15,13	23,66	7,15	0,47	0,855
○ Preventive benefit	52,38	19,89	54,69	11,67	-2,31	0,498
○ Therapeutic benefit	47,62	19,89	45,31	11,67	2,31	0,498
III.- Knowledge about the intervention	**18,98**	**12,58**	**19,46**	**6,82**	**-0,48**	**0,825**
• Quality of the evidence	59,03	21,23	61,57	10,34	-2,54	0,483
○Validity	54,89	18,64	51,34	11,33	3,55	0,267
○Relevance	45,11	18,64	48,66	11,33	-3,55	0,267
• Expert consensus	40,97	21,23	38,43	10,34	2,54	0,483
IV.- Economic impact	**13,92**	**9,98**	**18,54**	**7,35**	**-4,62**	**0,007****
• Direct healthcare costs	27,99	14,31	31,86	10,37	-3,87	0,117
• Other healthcare costs	25,24	13,58	21,17	5,98	4,07	0,9
• Non-medical costs	24,47	14,84	18,51	6,36	5,96	0,019*
• Opportunity cost and budgetary impact	22,30	14,56	28,46	11,94	-6,16	0,015*
V.- Feasibility	**14,46**	**10,43**	**15,17**	**5,19**	**-0,71**	**0,687**
• Availability of resources in the system	53,87	20,21	58,66	13,58	-4,79	0,169
• Organizational impact	46,13	20,21	41,34	13,58	4,79	0,169

Source: Own elaboration. p-values corresponding to the t-test.

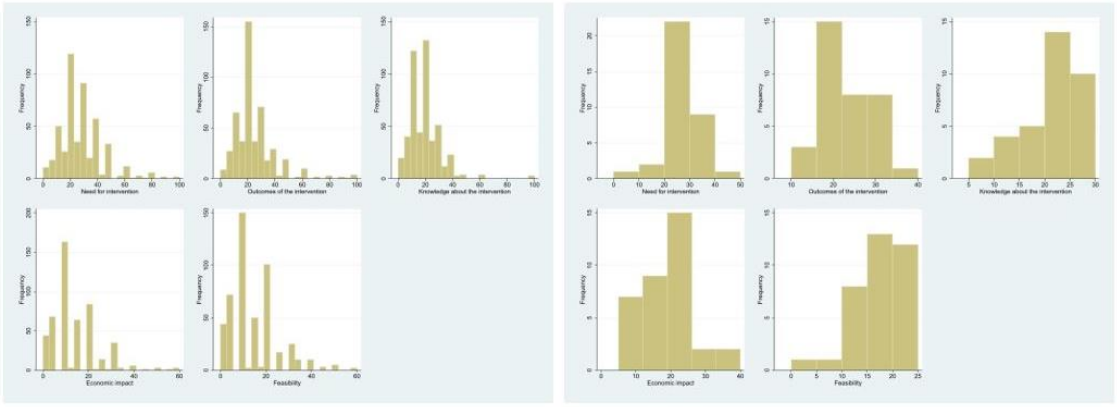
The average weight assigned by the general population is higher than that given by healthcare professionals for the first two domains, and lower for the remaining three domains. However, statistically significant differences (at the 95 percent confidence level) in mean weights between the two subsamples are observed only in the domain "Economic impact" (p=0.007).

Regarding the criteria, nine of them receive higher weights from the general population than from healthcare professionals, while six receive lower weights. Nevertheless, statistically significant differences are found only in one criterion of the first domain ("Affected population") and in two criteria of domain IV ("Non-medical costs" and "Opportunity costs and budget impact"). Lastly, none of the fifteen sub-criteria exhibit significantly different weights between the means of the two subsamples.[2]

The analysis of the distribution of absolute frequencies from the combined sample set (N=529), suggests a greater dispersion of scores in the first two domains compared to the rest, particularly the last two domains. The medians of the scores decrease as one progresses through the domains. The median for the “Need for intervention” domain is 25, followed by 20 for the “Outcomes of the intervention” and “Knowledge of the intervention” domains, and finally 10 for the “Economic impact” and “Feasibility” domains.

Differentiating between the two samples, histograms in Figure 1 confirm the higher concentration of weights assigned by the sample of health professionals within a narrower range, typically not exceeding 30, compared to the general population sample, which exhibits a more skewed distribution spreading to the right.

Figure 1. Histograms of the domains’ weights from each subsample

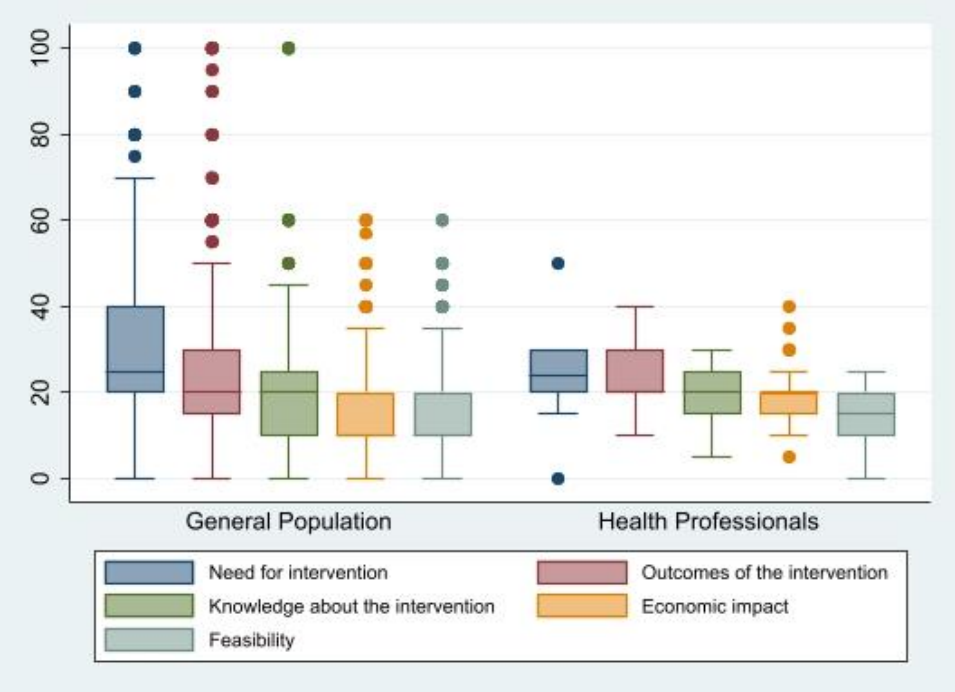


General Population (n=494) Healthcare professionals (n=35)

[2] When the Mann-Whitney-Wilcoxon test is employed, the statistically significant differences extend to the sub-criteria Impact on HRQoL (p = 0.034), Impact on life expectancy (p = 0.034), Change in prevalence (p = 0.043), and Unmet needs in effectiveness (p = 0.043).

In Figure 2 it is evident that the dispersion is significantly higher in the general population sample, although the medians, with the exception of the "Economic Impact" domain and, to a lesser extent, "Feasibility" are very similar. This greater homogeneity of the responses from the health professionals sample extends broadly when comparing the scores assigned to the criteria and sub-criteria, as shown in Table 2.

Figure 2. Weights assigned to the domains by each subsample



The different nature of the preferences and the significant difference in sample sizes between the two surveys make it impractical to integrate them into a single population to derive a measure of central tendency for establishing the weights. Combining the two samples would inevitably introduce bias towards social preferences, as they represent more than ninety-three percent of the total respondents. Therefore, we propose taking the average of the means obtained in the two samples for each item, that is, an equal-weighted sum of the mean weights from each subsample. By doing so, the resulting weights offer a more appropriate synthesis of both perspectives. These weights, rounded to the nearest integer, are presented in Table 3.

Once the high-impact technology has been valued, by assigning a score to each of the criteria and subcriteria -which falls outside the scope of this article, these scores should be combined with the weights in Table 3 as follows:

$$V = \sum_{i=1}^{5} \sum_{j=1}^{15} \sum_{k=1}^{15} \{(w_i S_i) \cdot (w_j S_j) \cdot (w_k S_k)\}$$

In the formula, i, j, and k represent the domains, criteria, and subcriteria of the analysis framework, respectively. The weightings from Table 2 are denoted as w_i, w_j, and w_k, representing the weights normalized to a total of one. S_i, S_j, and $S_k$represent the scores assigned by the decision makers to each domain, criterion, and subcriterion of the respective technology being evaluated.

Table 3. Weights (%) of domains, criteria and subcriteria for the MCDA.

Domain	**%**	**Criterion**	**%**	**Subcriterion**	**%**
Need for intervention	26	Disease severity	39	Impact on HRQoL	58
				Impact on life expectancy	42
		Affected population	34	---	
		Unmet needs	27	In effectiveness	30
				In HRQoL	27
				In safety	23
				In convenience	20
Outcomes of the intervention	24	Comparative effectiveness	27	Change in life expectancy	39
				Change in intermediate results	30
				Change in prevalence	31
		Comparative safety	27	---	
		Comparative patient-reported outcomes	24	Change in HRQoL	61
				Change in convenience	39
		Type of Benefit	24	Preventive benefit	54
				Therapeutic benefit	46
Knowledge about the intervention	19	Quality of evidence	60	Validity	53
				Relevance	47
		Expert consensus	40	---	
Economic impact	16	Direct healthcare costs	30	---	
		Other healthcare costs	23	---	
		Non-medical costs	22	---	
		Opportunity cost and budget impact	25	---	
Feasibility	15	Availability of resources in the system	56	---	
		Organizational impact	44	---	

Source: Own elaboration. The weights have been calculated as the average of the means of the two sub-samples.

Discussion

This article develops an MCDA framework for the evaluation of high-impact health technologies in a Spanish Regional Health Service. A multi-attribute function capable of generating a composite measure to assess the benefits and costs of high-impact health interventions is developed. The selection of the criteria was carried out in two stages. The authors made a pre-selection, based on the EVIDEM and KCE frameworks, followed by validation and final selection of the criteria by a group of decision-makers from the regional health system. The criteria were then weighted by two samples, one composed of decision makers and healthcare professionals and the other drawn from the general population.

Out of the five domains, "Need for intervention" and "Outcomes of the intervention" are the most highly weighted, both by the general population and the healthcare professionals' subsamples. "Affected population", "Disease severity", and "Quality of the evidence" ranked at the top among the 15 criteria, a result which is in line with other studies (Castro et al., 2016; Iskrov et al., 2016; Mirelman et al., 2012). While it is true that the first two mentioned domains absorb a 50% of the total value of the weighting function, the results of the weighting process also suggest that participants exhibit a certain tendency to distribute points equally between criteria and between sub-criteria. This pattern resembles, in some respect, the so-called equalizing bias (i.e., the tendency of decision-makers to assign the same weight to different attributes), which seems to affect particularly in point allocation rules, though the bias is less acute under a hierarchical structuring of the decision problem, such as the format used in our study (Rezaei et al., 2022). There seems to be also a tendency to use round numbers, which is common in this type of point allocation exercise (Honda et al., 2022).

Although a remarkable coincidence exists between the weights from the general population and those from the decision makers, some differences arise. First, healthcare professionals give more importance to the economic aspects of the intervention, which coincides with the results of previous studies in Spain (Caro et al., 2022) and in other countries (Claxton et al., 2015). The decision makers assigned to the domain "Economic impact" a weight which is more than 4.5

points higher than the weight derived from the general population's preferences. This could be explained by the fact that professionals are more aware of the budget constraint and, consequently, more sensitive to the costs of interventions and their economic impact in general.

Another interesting finding is that the dispersion of the weights of the domains is significantly higher in the general population sample than among the decision makers, although the medians turned to be very similar, with the exception of the "Economic Impact". This lower degree of dispersion of the responses provided by healthcare professionals seems a logical result, given that, firstly, the shared characteristics among members of this sample (employment status, level of education), as well as the presumably narrower age range it contains, make it more homogeneous. Secondly, it can be assumed that professionals may have more solidly formed opinions, and are therefore less prone to variability. Added to this is the disparate sample size of both groups of respondents, which may also help explain the differences in the degree of dispersion of the responses.

A controversial methodological issue has to do with the inclusion of cost-related attributes among the criteria. There are theoretical arguments for and against (Marsh et al., 2016). It has been argued that the aim of MCDA is to create a composite score of benefit, being the main question to be answered how much money should be spent for one unit of that composite score (Claxton 2015). Some researchers considered as unrealistic to assume that individuals are able to derive value functions for all criteria including costs and provide weights for the value function of costs in relation to that of the other criteria (Baltusen et al., 2019). Regarding cost-effectiveness, specifically, it has been recommended not to include it, from a technical perspective, since it is already a composite of costs and benefits (Wahlster et al.2015). One could assume, even, that the cost-effectiveness criterion, in some way, is implicitly included within the 'intervention outcomes' domain (Porter et al., 2010).

On the other hand, advocates of including costs argue that, by doing so, respondents explicitly make trade-offs between costs and the rest of the criteria, making explicit their contribution throughout the entire decision-making process (Rutten-Van Mölken et al.,2018). In a review of MCDA studies to support health technology assessment (Baltussen et al., 2019), eighty percent

of the studies included costs, and fifty-seven percent included cost-effectiveness, as criteria in the value measurement model. Another systematic review of criteria and scoring functions (Zelei et al., 2021) found that cost-related criteria were considered in more than fifty percent of the selected studies. In our study, we opted for including cost-related criteria in the MCDA framework, as it is the case in some recent studies (Caro et al., 2022; Claxton, 2015;Porter, 2010; Rutten-Van Mölken et al., 2018; Zelei et al., 2021;Campolina et al.,2022).

Incorporating the perspectives of various stakeholders is a fundamental aspect of MCDA. Stakeholder engagement ensures that the evaluation process reflects the values, concerns, and preferences of patients, healthcare professionals, payers, and policymakers. By involving stakeholders, MCDA fosters transparency, legitimacy, and acceptance of the final decision. Our study, as the Belgian framework (Cleemput et al., 2014), and in contrast to most examples in literature, incorporates the general population in the weighting stage, which is in line with the purpose of the MCDA scheme that has been designed, i.e., the incorporation of high-impact technologies into the public system. We think this is one of the strengths of the study, although we acknowledge as a potential limitation of the design the omission of incorporating the perspective of the general population (or the patients' perspective) in the initial phase of criterion identification.

Despite its advantages, MCDA faces certain challenges and limitations, and our study is no stranger to these. The selection and weighting of criteria can be subjective, leading to potential biases in decision outcomes, and this could be somehow present in our results. Particularly, the method chosen for weighting the criteria, namely, the 100-points allocation procedure, has been regarded as a more prone to framing bias, as criteria and their performance ranges are not explicitly traded off (Baltussen et al., 2019). Nevertheless, when choosing a method for weighting, time and resources required, as well as cognitive burden imposed to participants should also be considered (De Montis et al., 2004). The method we chose has the advantage of its simplicity and understandability, and it has been successfully used in previous studies (Zozaya et al., 2022).

On the other hand, the advisability of incorporating a deliberative component into any quantitative MCDA has been suggested (Baltussen et al., 2019), allowing the decision-making body to carry out a flexible interpretation of the results. This is the spirit that guides the proposal, not that of providing a rigid framework where the score obtained with the multi-attribute function becomes the sole input to consider in the decision-making process.

Finally, validation of the proposed framework would require its application in order to detect possible shortcomings or dysfunctions that could become apparent at the time of its use for the evaluation of a specific intervention or technology. The availability and reliability of data for all criteria could pose practical difficulties. And furthermore, interpreting and communicating the results of MCDA to diverse stakeholders can be complex, demanding effective communication strategies.

Future research, afterwards the framework has been used for a time, could check whether it has indeed been useful for decision-makers of the regional health service. A reassessment of its suitability should be done periodically and, depending on its success for making better decisions, to transfer to other instances.

Conclusions

Multi-Criteria Decision Analysis constitutes a valuable approach to systematically and transparently support decision-making, enabling a comprehensive evaluation of healthcare technologies based on various criteria. This article presents a multi-criteria decision scheme to guide the purchasing decisions of new high-impact technologies in a Spanish regional health service where, currently, no formal procedure with objective criteria exists for adopting such decisions. The development of the scheme has taken into account, in its different phases, the preferences of managers, healthcare professionals, and the general population. Although the contributions of the former have shown a higher degree of consistency and lower dispersion than the preferences of the general population, no significant discrepancies have been detected in how criteria are prioritized between the two groups. The result is a multi-attribute function capable of generating a composite measure to assess the costs and benefits of high-impact

interventions, with ‘need for intervention’ and ‘outcomes of the intervention’ emerging as the most relevant domains or attributes. The application of this framework in a specific decision context would provide valuable information about the effectiveness of this tool in informing priority setting in resource allocation within the regional health system.

Appendix 1

Table A 1. Composition of the discussion group responsible for selecting domains and criteria

	Gender (M/W)
Director Manager of the Regional Health Service ('Servicio Murciano de Salud': SMS)	M
Director General of Healthcare of the SMS	W
Deputy Director of Healthcare Quality, Safety, and Evaluation of the SMS	M
Deputy General Director of Economic Affairs of the SMS	M
Deputy General Director of Projects and Innovation of the SMS	M
Director General of Hospital Care of the SMS	W
Head of Service of the Health Service (Management of Care Coordination Programs)	W
Director Manager of Health Area 1 and Virgen de la Arrixaca Hospital	M
Director Manager of Health Area 2 and Santa Lucía Hospital	M
Director Manager of Health Area 6 and Morales Meseguer Hospital	M
Head of the Comprehensive Supply Unit (Procurement Center) of the SMS	M

Table A 2. Sociodemographic characteristics of the general population sample

Age and sex							
	Men		**Women**		**Total**		**Region of Murcia (1)**
	N	**%**	**N**	**%**	**N**	**%**	**%**
18 – 29 years	49	19.9	45	18.1	94	19.0	17.2
30 – 44 years	69	28.0	67	27.0	136	27.5	26.6
45 – 60 years	65	26.4	64	25.8	129	26.1	30.2
> 60 years	63	25.6	72	29.0	135	27.3	26.0
Total	246	100.0	248	100.0	494	100.0	100.0
Average Age (desv.st)	47 (16.9)						48.6
% by sex	49.8		50.2		100.0		49.8 / 50.2
Educational level							
	N		**%**		**Accumulated Percentage**		**Region of Murcia (%)**
Without studies	20		4.1		4.1		8.8
Primary	93		18.8		22.9		8.3
Secondary	219		44.3		67.2		57.6
Superior	162		32.8		100.0		25.3
Civil status							
	N		**%**		**Accumulated Percentage**		**Region of Murcia (%)**
Single	177		35.8		35.8		46.9
Married	228		46.2		82.0		42.0
Divorced	57		11.5		93.5		5.9
Widow/er	32		6.5		100.0		5.3
Occupation							
	N		**%**		**Accumulated Percentage**		**Region of Murcia (%)**
House husband/wife	29		5.9		5.9		10.9
Student	41		8.3		14.2		9.9
Retired	92		18.6		32.8		15.1
Own account	68		13.7		46.6		9.4
Employed	201		40.7		87.3		46.2
Unemployed	63		12.8		100.0		8.5

(1) Spanish National Institute of Statistics: Census and Labor Force Survey. 2022.

Appendix 2. Glossary of domains, criteria and subcriteria

Disease severity: severity of the health condition of the patients treated with the intervention (or severity of the condition that is intended to be prevented) in relation to mortality, morbidity, disability, functioning, impact on quality of life, disease course (intensity, clinical stages). The impact of disease severity on health-related quality of life (Impact on HRQoL), as perceived by patients, as well as on survival (Impact on life expectancy) is considered.

Affected population: number of people affected by the condition (treated or prevented with the intervention) among a specific population in a given period of time. It can be expressed as the annual number of new cases (incidence) and/or as the proportion of the population affected at a specific moment in time (prevalence).

Unmet needs: deficiencies of the interventions being compared in their capacity to prevent, cure or mitigate the health problem to which they are addressed (Effectiveness). Included among these deficiencies are those corresponding to safety (Safety), the results perceived by patients (HRQoL) and convenience or comfort (Convenience).

Comparative effectiveness: capacity of the intervention to prevent or produce a desired (beneficial) change in the symptoms or in the course of the condition superior to the beneficial changes generated by alternative interventions. This change is captured in the form of increased life expectancy (Life expectancy), better intermediate or surrogate results (Intermediate results) and prevalence (Prevalence).

Comparative safety: ability of the intervention to produce a reduction in unwanted or harmful effects related to the intervention compared to those caused by alternative interventions (Adverse events).

Patient-reported outcomes: ability of the intervention to produce beneficial changes in patient-perceived outcomes greater than the beneficial changes produced by alternative interventions. It includes changes perceived by patients in terms of health-related quality of life (HRQL) as well as convenience or ease of use (Convenience).

Type of benefit: if the intervention yields a preventive benefit or risk reduction (eg eradication, prevention, reduction in transmission, reduction in the prevalence of risk factors) or a therapeutic benefit (eg relief of symptoms, prolongation of survival, cure).

Quality of the evidence: extent to which the available evidence on the intervention is relevant (*Relevance*) for decision-making (in terms of population, disease progression, comparators, outcomes, etc.) and valid (*Validity*) with respect to scientific standards (study design) and previous conclusions (degree of concordance of the results obtained between different studies). Includes considerations about the uncertainty of the evidence (e.g. conflicting results between studies or a limited number of studies and patients). Having a complete report of the available evidence is a prerequisite to assess its coherence and validity.

Expert consensus: degree of suitability of the intervention (or of similar alternatives) according to the existing consensus among experts about what constitutes appropriate medical practices for the management of the health condition in question. The recommendations of the experts in this regard are usually made explicit in the form of clinical practice guidelines with the intention of improving the quality of this.

Direct healthcare costs: net cost of financing the intervention, understood as the difference between the expected cost of the intervention and that corresponding to other interventions that could be replaced by it. This comparison is limited to the cost directly linked to the intervention, that is, the cost of acquisition, implementation and maintenance.

Other health costs: impact of the intervention on other health costs (excluding the direct cost of the intervention), such as hospitalization costs, specialized consultations, those caused by possible adverse effects, long-term care, etc.

Non-medical costs: impact of the intervention on non-health costs, such as those corresponding to social services, productivity losses, informal care, etc.

Opportunity cost and budget impact: consideration of the health resources that may be sacrificed (opportunity cost) as a result of the implementation of the intervention and the system's capacity to assimilate such intervention budgetarily.

Availability of resources in the system: material and, above all, human resources available to the system to guarantee a priori the implementation of the intervention, without the need to displace other services or programs in force. The displacement, where appropriate, of these would be included in the preceding criterion ("Opportunity cost and budgetary impact").

Organizational impact: magnitude of the reorganization of available resources that must be undertaken, if applicable, in order to address the considered intervention.

Appendix 3. Some examples of the weighting procedure

Figure A 1. Example of distribution of the 100 points among the 5 domains.

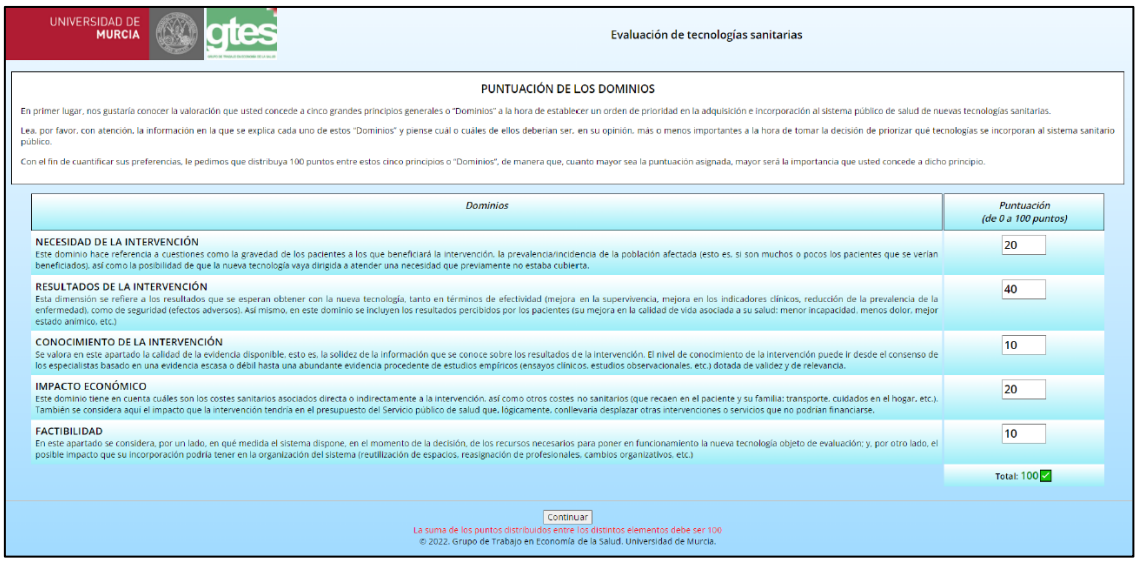
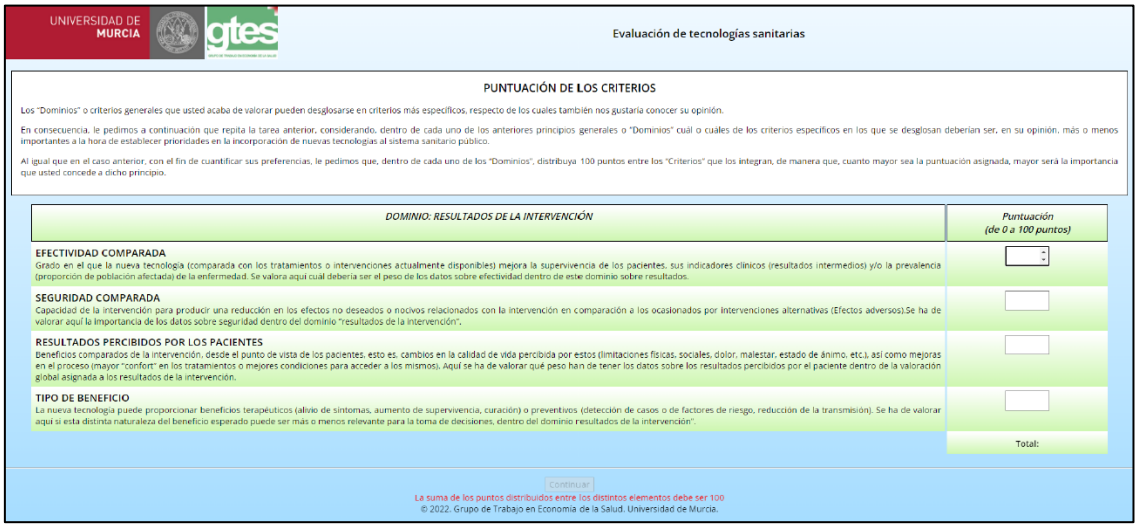


Figure A 2. Weighting task of the criteria of the "Results of the intervention" domain.



3.THE INTRINSIC VALUE OF THE INFORMATION CONTAINED IN MEDICINE LEAFLETS[3]

Abstract

Aim. The aim of this paper is to estimate the monetary value of the information contained in medicine leaflets applying contingent valuation. **Method.** By surveying a sample of the general population, we obtain willingness-to-pay estimates of the value of providing additional quantitative information on potential benefits and side effects of a hypothetical medicine, according to the best evidence available about risk communication. **Results.** The willingness-to-pay estimates found in our study ranged from 60 cents to 1 euro per month. In addition, some consistency tests of the robustness of our estimates are also presented, as well as evidence on their feasibility, reliability and validity. **Conclusions.** This paper provides compelling evidence supporting the notion that access to information about the effectiveness and side effects of drugs holds inherent value for the population.

Keywords: Contingent valuation, willingness to pay, patient information leaflets, risk communication.

[3] This chapter has been published in *Hacienda Pública Española/Review of Public Economics*. https://dx.doi.org/10.7866/HPE-RPE.24.2.4

Introduction

Standard economic theory (and health economics as well) conventionally assumes that the final outcomes of the different alternatives are the sole determinants of value for individuals. Consequently, benefits of health interventions are usually identified only with the health gains they produce.

Nevertheless, health interventions may also yield utility for patients because of what Donaldson and Shackley (1997) call "process" attributes (i.e., attributes not strictly related to health but that impact individual well-being as well). One example would be better convenience of a medication regimen. Another could be the value of patient information leaflets that accompany prescription drugs. Precisely, this is the main aim addressed in this paper: to estimate the intrinsic value that people attribute to information about medication contained in the leaflets.

As is well known, the contingent valuation (CV) methodology is a survey-based approach rooted in applied welfare economics used to assign a monetary value to non-market goods such as the one (i.e., value of patient information leaflet) we are interested in this paper (Mitchell and Carson, 1989). Typically, that value is approximated by eliciting willingness to pay (WTP) or willingness to accept (WTA) estimates from respondents. Although this method has been extensively used to evaluate health treatments and, moreover, to evaluate how levels of information impact WTP for health programs (e.g., Donaldson & Shackley, 1997; Pinto et al., 1998; Protiere et al., 2004; Shono et al., 2014; Dealy et al., 2017), there is just one study to date (Dealy et al., 2021) that has attempted to estimate the WTP for the information contained in the leaflets. More specifically, the authors estimated the WTP for changing from non-standardized formats of providing information on uses, side effects and interactions of a medicine to a standardized format.

This paper builds on Dealy et al. (2021) work by going one step further. We also compare two different formats. One of them represents a "traditional" leaflet, similar to the non-standardized formats used by Dealy et al. As in those used by these authors, no quantitative data about

efficacy and side effects is provided in our traditional leaflet. The other format employed in our study, however, provides additional quantitative information about potential benefits and harms of the same hypothetical medicine presented in the traditional format. Moreover, that information is given according to the best evidence available about risk communication, depicting it in the form of natural frequencies through a visual aid.

In this respect, there is evidence suggesting that the readability of patient information leaflets (PILs) is frequently too complex for average patients to comprehend fully (Hamrosi et al., 2012). Simplifying language, using natural frequencies rather than percentages for probabilities, and consistency in risk expressions can enhance comprehensibility (Gigerenzer et al., 2007). The framing and formatting of risk information impacts perception. Patients often overestimate the significance of risks when provided relative measures such as "twice as likely" compared to absolute risks or natural frequencies (Sirota & Juanchich, 2019). Nevertheless, presenting risks in the form of frequencies of type, for example, 1/10,000, may be misleading, because such information may trigger different impacts depending on equivalent but superficially different methods of presentation (Pinto et al., 2006). So, to avoid this bias, as Yamagishi (1997) suggests, the specific risk about which we want to inform patients should be presented in comparison to other risks (e.g., a heart attack risk could be presented as a comparison in frequencies of premature deaths due to traffic accident, diagnostic X-rays or tobacco smoking) following a descending order. Furthermore, visual aids are also demonstrated to facilitate patient understanding of risk magnitudes. Pictographs, icon arrays, and bar charts can allow quicker interpretation of probability data than textual descriptions or percentages alone (Garcia-Retamero & Galesic, 2011). Intermittent bolding of key risk descriptors can help draw attention to important details and visual presentations are also optimally interpreted when arranged in decreasing order of risk magnitude rather than randomly (Sirota et al., 2018). Besides, providing balanced information about side effects and treatment benefits can avoid skewing risk perceptions (Webster et al., 2018). Tailoring visual aids and risk communication approaches to

suit health literacy levels and norms in target populations enhances effectiveness (Yi et al., 2015).

On the basis of the previously discussed evidence-based principles for effective risk communication, this paper reports WTP estimates of the value provided by a Spanish general population sample to the information conveyed in a way aligned with such principles through a complementary brochure, additional to the traditional PIL included in the package of a hypothetical medicine. In addition, some consistency tests of the robustness of our estimates are also presented, as well as evidence on their feasibility, reliability and validity.

The manuscript is organized as follows. First, elicitation methods, survey design, and questionnaires used are described. Besides, the main hypothesis to be tested, consistency tests and statistical methods employed are explained. Next, main findings are presented. The discussion closes the paper

Methods

General outline of the survey design

A contingent valuation (CV) study was designed to estimate the value of the additional information contained in a complementary brochure to the usual patient information leaflet (PIL) of a hypothetical medicine. This medicine was an anticoagulant indicated for preventing cardiovascular diseases. We chose this drug because it involves a clear trade-off between benefits (preventing heart attacks) and risks (increased susceptibility to gastrointestinal bleeding).

CV surveys elicit individuals' willingness to pay (WTP) using different elicitation formats. However, there are some concerns about biases distorting WTP responses, such as hypothetical biases (i.e., where stated value differs from true value) (Haab et al. 2013) and scope effects (i.e., insensitivity of WTP responses to changes in goods) (Carson, 1997). Aside from these potential biases, which are common to any CV study, applications in the specific context of healthcare

present an additional challenge compared to other fields. This challenge arises from the fact that respondents may be reluctant to put "a price" on health, given that it may be perceived as a superior good.

Addressing all these issues requires making a series of careful decisions concerning questionnaire design. In this way, a realistic and comprehensible description of the hypothetical scenario recreated in the survey is necessary. Likewise, a suitable payment mechanism (e.g., price, tax, insurance premium) has to be chosen, as well as the elicitation format (e.g., open-ended, close-ended, payment card) through which WTP responses will be elicited, among other factors. Additionally, it is essential to incorporate supplementary questions into the questionnaire to conduct tests validating the results. Next, all these decisions, applied to our study, are described.

The sample

The sample comprised 217 adults selected to represent the age and gender distribution within the general population. They were randomly assigned to two distinct groups (n1 = 110; n2 = 107). The questionnaires administered to the subjects in each of the groups were identical, except for the section inquiring about their willingness to pay (WTP). The survey took place in the region of Murcia, Spain, over a period of 1 month. All the interviews were face-to-face, and the average time per interview was around 20 minutes. A subsample of 20 individuals responded to the questionnaire a second time, 15 days after their initial interview, to evaluate the test-retest reliability of their responses.

The questionnaire

The questionnaire was structured in four parts. First, the study was introduced to the participants. Next, throughout the second part of the questionnaire, the nature of myocardial infarction—the disease that the anticoagulant medication is intended to prevent— and the consequences of gastrointestinal bleeding associated with that medication were explained to participants with the help of the visual aids provided in the Appendix.

Part 3 of the questionnaire contained WTP questions. Prior to asking these questions, two different, though complementary, pieces of information on the anticoagulant medicine and its consequences were presented to the participants. The first piece of information (Figure 7) was a "traditional" leaflet, and the latter one (Figure 8) a brochure providing additional quantitative information on the effectiveness and side effects of the medicine. As noted in the Introduction, the design of this complementary brochure responds to best practices on risk communication reported in the literature, particularly Yamagishi's (1997) recommendations for effective risk communication.

Once both the PIL and the complementary brochure were shown to the respondents, they were asked to state their maximum WTP for the medicine in different ways depending on the group to which they belonged, as explained further in the next section. The payment scenarios involved a monthly payment for a year. To make the WTP questions plausible and avoid protest responses (i.e., stating a zero bid or refusing to state a bid), we strived to justify the absence of public financial coverage for the drug by emphasizing its preventive nature. So, the participants were informed that there was another medicine funded by the public system specifically for people who have heart disease. Additionally, interviewers were briefed on the possibility of encountering these types of responses and their role in making the scenario credible, as well as addressing any initial reluctance some respondents may have.

The last part of the questionnaire (Part 4) included a set of questions aimed firstly at gathering sociodemographic information (i.e., gender, age, marital status, educational attainment, employment status, and income level). Some of these variables were included to examine potential effects on the responses. Notably, income level values were used to assess whether WTP estimates increased with income, one of the tests that CV results have to pass to claim their construct (Carson and Mitchell, 1989) or theoretical validity (Bateman et al., 2002), i.e., that WTP responses agree with expectations predicted by economic theory. Some other questions also inquired about the participants' previous experience with health problems such as heart attacks and gastrointestinal bleeding. Finally, a series of questions gauged the

interviewee's comprehension of essential questions related to the health issues under consideration and assessed their numerical skills, specifically their ability to convert frequencies expressed in different formats. The specific ability to compute proportions, percentages or probabilities is often referred to as statistical numeracy (Cokely et al. 2014), being one of the drivers of risk literacy (i.e., the ability to accurately interpret and act on information about risk). So, the inclusion of these questions in our CV survey allowed us to check out whether statistical innumeracy influenced WTP estimates.

Responses to all these questions were incorporated as independent variables into the different econometric models used to examine the influence of sociodemographic, economic, and personal traits as potential determinants of the estimated WTP for the information and the medicine. Likewise, as noted above, it also served to check whether WTP estimates conform to the predictions of economic theory.

Willingness to pay questions: elicitation procedure

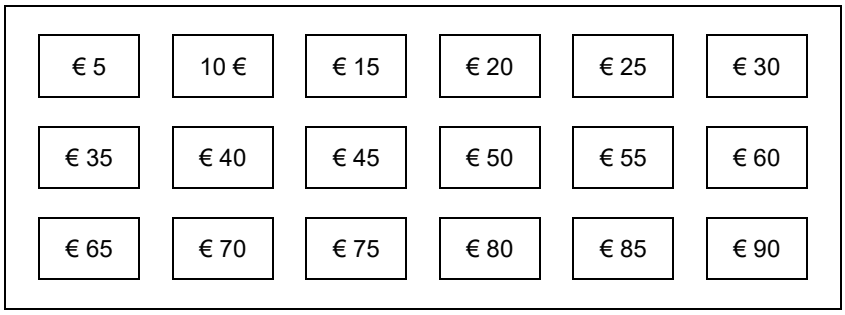
As mentioned above, the questionnaires administered to the subjects in each of the groups were identical, except for how they were asked about their WTP. In Group 1, participants first stated their WTP for the medicine whose package only included the traditional leaflet, and then stated their WTP for also including the brochure containing detailed information on the anticoagulant's benefits and harms within the package. Thus, they valued the medicine and the complementary information provided by the brochure separately. On the contrary, in Group 2, participants stated their maximum WTP for the medicine package as a whole, including both the PIL and the complementary brochure. Afterwards, respondents belonging to Group 2 were asked to indicate the proportion (as a percentage) of the total WTP they had previously stated that was attributed exclusively to the additional information provided in the brochure.

This different procedure to estimate the value of the information served to test the consistency of the results across the two groups in a twofold sense. On the one hand, it allowed us to test if the different response format used in each group (i.e., a money amount in Group 1 vs a

percentage over the overall WTP for the whole medicine package) to elicit the values for information led to a significant discrepancy between their WTP estimates. On the other hand, it made possible to test whether a 'part-whole' effect (Bateman et al., 2006, 2007) occurs, i.e., whether the sum of the WTP for the medicine and the WTP for the brochure, when valued separately as they were done in Group 1, was larger than the WTP placed on the medicine package including the brochure as a whole, such as was performed in Group 2.

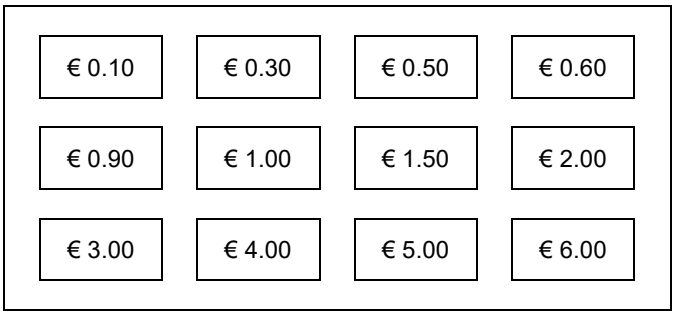
To elicit WTP values a mixed format was adopted, combining a payment card method with an open-ended question. First, respondents indicated if they would pay or not each of the different amounts shown on the payment card (Figures 3 and 4) to narrow the range of possible WTP values, and then they specified their definitive WTP through an open-ended question. This format, first used by Carthy et al. (1999), has also been employed by Sánchez-Martínez et al. (2021) to elicit the value of a statistical life, showing a reasonable balance between preference elicitation feasibility and accuracy of responses.

Figure 3. Payment card for the valuation of the medicine (*)



(*) To evaluate the medicine in the strict sense in sub-sample 1 and for the medicine and the additional information in Group 2.

Figure 4. Payment card for the valuation of the brochure (Group 1)



To facilitate participants in Group 2 stating the percentage of their WTP for the medicine package attributed to the complementary brochure included in the package, a visual analogue scale (VAS) resembling a graduated thermometer ranging from 0 to 100 was used.

Hypothesis testing and internal consistency

The central hypothesis we aim to test posits that individuals assign a positive value to the availability of information about the benefits and side effects of medicines. Besides testing this key hypothesis, in parallel we also check out the internal consistency of our findings or, as Kling et al. (2012) argue, the extent to which stated preferences are consistent with theoretical expectations (construct validity).

Therefore, we first test the following hypothesis:

> ***Main hypothesis****: The WTP for receiving quantitative information on the effectiveness and side effects of the medicine (the brochure) will be significantly greater than zero.*

This hypothesis will be supported if, denoting WTP $(i)^1$ as the WTP for the complementary brochure directly stated by participants in Group 1 and WTP $(i)^2$ as the WTP attributed to the same brochure in Group 2, the following conditions are met:

$$WTP\ (i)^1 > 0$$

$$WTP\ (i)^2 > 0$$

Since respondents in Group 2 are not directly asked to state their maximum WTP, but rather they are asked to set the percentage representing the value of the information regarding the

WTP for the total cost of the medicine, which includes the additional information, WTP $(i)^2$ is calculated as the product:

$$\% \text{ WTP } (i)^2 \times \text{ WTP } (m/i)^2,$$

Where % WTP $(i)^2$ denotes the percentage that represents the value of the information and WTP $(m/i)^2$ the WTP for the drug, including the information.

In addition, as noted above, our split-sample design allows us to test the internal consistency of our CV survey by addressing two potential anomalies in stated preferences. Firstly, we will examine whether the different response format used to elicit WTP values for the brochure in each group will lead in turn to different WTP estimates. Elicitation effects are recognized as being related to violations of the principle of procedure invariance (Tversky et al., 1988), whose image in CV studies is the persistent finding that the response format systematically affects reported WTP values (Poe, 2016). In our case, testing for elicitation effects implies to compare the WTP for the information elicited in isolation from Group 1 to the WTP for the information inferred from applying the percentage attributed by respondents in Group 2 to the stated WTP for the whole package. In consequence, if elicitation effects are present in our data then the value assigned to the brochure will not be independent on the response format, which implies:

$$\text{WTP } (i)^1 \neq \text{WTP } (i)^2$$

Which means in turn that:

$$\text{WTP } (i)^1 \neq \% \text{ WTP } (i)^2 \times \text{ WTP } (m/i)^2$$

Lastly, we will examine whether there are significant differences between, on the one hand, the sum of the WTP elicited for the medicine and the complementary brochure separately (Group 1) and, on the other hand, the overall WTP for the entire medicine package including the brochure (Group 2). So, we propose to test for 'part-whole' effects, one of the anomalies involved into the so-called embedding effects (Kahneman and Knetsch, 1992). Within embedding effects there is room for ordering effects, scope insensitivity, visible choice-set effects and part-whole/substitution effects (Bateman et al., 2006, 2007). The latter anomaly occurs in the context

of CV studies when it appears that the sum of the valuations placed by respondents on the parts of a good is larger than the valuation placed on the good as a whole. Transferring this to our survey, if there are part-whole effects in it then:

$$WTP\ (m)^1 + WTP\ (i)^1 > WTP\ (m/i)^2,$$

where $WTP\ (m)^1$ represents the WTP for the medicine elicited in Group 1.

Statistical methods

To test the key hypothesis of the paper, i.e., whether WTP for the complementary brochure is significantly greater than zero, both the parametric one-sample *t*-test and the non-parametric one-sample Wilcoxon signed-rank test will be used. Likewise, we will try to identify the main determinants of the WTP for the complementary brochure by means of a two-step econometric strategy. Firstly, an ordinary least squares (OLS) regression analysis, using different specifications, will be performed. Percentages attributed to the value of the additional information contained in the brochure will be converted to euros as explained above, by multiplying the proportion of the overall WTP for the whole medicine package (including the brochure) that respondents assign to the additional information times that overall WTP. Next, in case that WTP values are far from a normal distribution, a censored model (Tobit model) will be estimated.

Consistency of WTP responses will be also analysed by using both regression analysis and, in this case, two-sample statistical tests. In this way, to check whether the value of the additional information contained in the brochure is not independent of the elicitation format used (separate valuation vs. valuation "implicit" in the overall value for the medicine, including the brochure), a dichotomous variable capturing whether the respondent answered questionnaire 1 or 2 is included as a regressor (see Table 4). Additionally, WTP estimates obtained in both formats (WTP for the complementary brochure, in the first case, and the percentage attributed to the additional information multiplied by the WTP for the whole medicine package, in the second

one) are compared using both the unpaired *t*-test and the non-parametric Wilcoxon-Mann-Whitney (WMW) test.

Likewise, testing for 'part-whole' effects is performed through another regression analysis, including newly the questionnaire type as an independent variable. To ensure homogeneity between the observations from both groups, estimates used for Group 1 in the analysis are those obtained by combining the two types of WTP values (i.e., for the medicine itself and for the additional information). Conversely, for Group 2, the WTP for the whole medicine package (including the complementary brochure) is used. This analysis is supplemented by comparing WTP estimates elicited from Group 2 to the sum of the two types of WTP estimates (i.e., for the medicine and the complementary brochure) elicited from Group 1 using both parametric and non-parametric tests.

Table 4. Brief description of the regressors used.

	Description
Sex	Dichotomous variable. Male (1); Female (0)
Married	Dichotomous variable. Married (1); Otherwise (0)
Primary studies	Dichotomous variable. Primary studies (1); Otherwise (0)
Secondary studies	Dichotomous variable. Secondary studies (1); Otherwise (0)
Higher education	Dichotomous variable. Higher studies (1); Otherwise (0)
Age	Continuous variable
Age^2	Variable *Age* squared
35<=Age<50	Dichotomous variable. 35<=Age<50 (1); Otherwise (0)
50<=Age<65	Dichotomous variable. 50<=Age<65 (1); Otherwise (0)
Age>=65	Dichotomous variable. Age>=65 (1); Otherwise (0)
Income	Continuous variable
Heart attack experience	Dichotomous variable. Has had experience (1), Otherwise (0)
Digestive bleeding experience	Dichotomous variable. Has had experience (1), Otherwise (0)
Numerical skills	Dichotomous variable. Mathematical skill (1), Otherwise (0)
Questionnaire Type 2	Dichotomous variable. Group 2 (1), Otherwise (0)

Table 4 provides a brief description of all the explanatory variables included in the different econometric specifications that will be considered. The first group of variables corresponds to

the sociodemographic characteristics of the subject: gender, marital status, age (including its square, considering the possibility of non-linear effects), and level of completed education. Age is also included as a discrete variable in order to test a potential "pensioner effect" for the WTP for the medicine. Such an effect is the label given to describe the reluctance that population aged 65 and over (mostly pensioners) can feel toward paying a higher WTP, since what they usually pay for medicines, due to the co-payment structure in Spain, is lower than working-age population do. The inclusion of income among the explanatory variables in the different regressions that will be performed will serve to test if WTP correlates well with income, so a positive and significant income coefficient is to be expected. Likewise, a variable has been included to capture the numerical skills of the respondent, specifically their ability to convert risks expressed in terms of 10,000 to risks expressed as percentages. Those who answered both questions of this type correctly in the survey were considered mathematically 'competent.' Additionally, two dichotomous variables accounting for prior experience with heart attacks and digestive bleedings, respectively, has been included, along with the variable capturing the group to which the respondent belongs.

Finally, a test-retest reliability analysis will also be conducted, for which Pearson and Spearman correlations will be calculated. A high Pearson correlation (close to 1) signifies strong linear agreement between initial and retest scores, while a high Spearman correlation indicates consistent monotonic agreement.

Results

Characteristics of the sample

Table 5 shows main features of the sample and the two groups than comprise it. Roughly speaking the sample is representative of the Spanish general population.

Table 5. Characteristics of the sample.

	Group 1		***Group 2***		***Total***	
	n	***%***	***n***	***%***	***n***	***%***
Age						
less tan 35 years	38	34.5	36	33.6	74	34.1
35 to 49 years	38	34.5	36	33.6	74	34.1
50 to 65 years	22	20.0	22	20.6	44	20.3
65 and older	12	10.9	13	12.1	25	11.5
Mean Age (Std. Dev.)	41.07 (15.2)		42.33 (15,8)		41.69 (15.47)	
Sex						
Man	54	49.09	52	48.6	106	48.8
Woman	56	50.9	55	51.4	111	51.2
Married						
Yes	62	56.4	62	57.9	124	57.1
No	48	43.6	45	42.1	93	42.9
Level of studies						
No studies	10	9.1	6	5.6	16	7.4
Primary studies	30	27.3	36	33.6	66	30.4
Secondary studies	48	43.6	44	41.1	92	42.4
Higher education	22	20.0	21	19.6	43	19.8
Monthly income						
Between 600 and 1200 euros	35	31.8	37	34.6	72	33.2
Between 1200 and 1800 euros	22	20.0	23	21.5	45	20.7
Between 1800 and 2,700 euros	39	35.5	40	37.4	79	36.4
More than 2700 euros	10	9.1	11	10.3	21	9.7
Mean Income (Std. Dev.)	1689.8 (716.52)		1748.90 (691.8)		1719.12 (704.44)	
Numerical skills						
Competent	35	31.8	42	39.3	77	35.48
No competent	75	68.2	65	60.7	140	64.52
Heart attack experience						
Yes	5	4.55	3	2.80	8	4.15
No	105	95.5	104	97.20	209	95.85
Digestive bleeding experience						
Yes	8	7.27	13	12.15	21	9.68
No	101	91.82	94	87.55	196	90.32
TOTALS	**110**		**107**		**217**	**100.0**

Willingness to pay estimates

In Group 1, all the participants stated a positive WTP value for the complementary brochure. However, the same did not happen for respondents in Group 2, where a total of 35 interviewees

assigned a zero WTP value for the complementary brochure medicine.[4] Nevertheless, as it is shown in Table 6, despite these zero values, mean WTP in Group 2 was positive and larger than that obtained in Group 1. Both parametric and non-parametric tests suggest that WTP estimates in the two groups are significantly different from zero (one-sample *t*-test and one-sample Wilcoxon signed-rank test, $p < 0.0001$), supporting in this way hypothesis 1.

Table 6. Willingness to pay for the complementary brochure (€).

	Group 1	Group 2	
	WTP(i)[1]	WTP(i)[2]	WTP(i)[2] (>0)
Mean	0.63	1.01	1.5
Median	0.50	1.00	1.375
Std Dev	0.766	1.156	1.115
Min.	0.1	0	0.25
Max.	6	6	6
N	110	107	72

WTP(i)[1]: WTP directly stated in euros. WTP(i)[2]: WTP converted to euros. Last column: participants in Group 2 with WTP(i)>0.

As noted in the Methods section, we conducted an OLS regression analysis of the determinants of the WTP for the complementary brochure. Different OLS estimations have been performed by changing the regressors in each of the models 1 to 3 in Table 7. Additionally, since WTP figures are far from a normal distribution (see Figure A5 in the Appendix 4), a censored model (Tobit model) has also been estimated. To perform these analyses, WTP estimates elicited from Group 2 were previously converted to euros, as indicated at the bottom of Table 6.

[4] Indeed, there were 34 participants who, having stating a positive WTP value for the whole medicine package (including the complementary brochure), then set a null percentage over that value onto the VAS. The remaining participant until amount the 35 respondents who did not provide any value to the complementary brochure was the only participant in the whole sample who was not willing to pay anything for the medicine.

Table 7. Results of the regressions of the WTP for the brochure, WTP(i).

	OLS estimations			Tobit Model
	(1)	(2)	(3)	
Sex	-0.210	-0.197	-0.101	-0.236
	(0.142)	(0.135)	(0.139)	(0.154)
Married	-0.245	-0.252	-0.142	-0.218
	(0.167)	(0.165)	(0.165)	(0.190)
Age	0.0586**	0.0586**	0.0875***	0.0574**
	(0.0237)	(0.0235)	(0.0216)	(0.0266)
Age2	-0.0008***	-0.0008***	-0.0011***	-0.00083***
	(0.00025)	(0.00025)	(0.00023)	(0.00029)
Heart attack experience	0.283	0.242	0.227	0.399
	(0.325)	(0.321)	(0.330)	(0.362)
Digestive bleeding experience	-0.0564	-0.0415	-0.0675	-0.0847
	(0.225)	(0.223)	(0.228)	(0.261)
Numerical skills	-0.0886	-0.142	-0.0645	-0.172
	(0.162)	(0.153)	(0.164)	(0.172)
Questionnaire Type 2	0.389**	0.377**	0.467***	0.166*
	(0.161)	(0.152)	(0.161)	(0.71)
Income	0.00035***	0.00031**		0.00038***
	(0.00013)	(0.00012)		(0.00014)
Primary studies	-0.223		-0.133	
	(0.285)		(0.288)	
Secondary studies	-0.154		0.0293	
	(0.341)		(0.340)	
Higher education	-0.380		-0.0923	
	(0.366)		(0.356)	
35<=Age<50				
50<=Age>65				
Age>=65				
Constant	-0.368	-0.511	-0.720	-0.487
	(0.526)	(0.438)	(0.518)	(0.493)
Observations	216	216	216	216
R-squared	0.204	0.196	0.175	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The figures in Table 7 suggest a positive relationship between age and WTP, and this relationship also appears to be nonlinear. The significance and negative sign of the Age^2 coefficient in models 1 to 3 indicate an inverted U-shaped relationship between WTP and age, similar to the age-wage profiles observed in the labour market. This suggests that the age variable may partially capture the income effect. The income variable also proves to be statistically significant, thereby confirming the theoretical validity of the questionnaire in this

respect, as we will discuss later. Educational levels are non-significant, even when, considering the possibility of a high correlation between education and income, the variable income is omitted (model 3). Other variables such as prior experience with heart attack and digestive bleeding, or the mathematical competence of the subjects turned out to be non-significant too. The results of the censored model are qualitatively similar to those of the OLS regressions.

Table 7 apparently indicates that the way in which WTP for additional information was obtained (as a separate WTP question in group 1 or as a proportion of the overall WTP for the medicine in questionnaire 2) affects WTP values. The positive and statistically significant coefficient of the 'Questionnaire type 2' variable in all models suggests that determine the WTP for the complementary brochure as a percentage of the overall WTP for the medicine makes that the value conferred to the additional information is higher, such as Table 8 indeed shows. Notwithstanding, though a significant difference between the WTP estimates provided in both groups is found with the unpaired *t*-test ($p = 0.0052$), significance is not reached according to the non-parametric WMW test ($p = 0.11$).

Table 8 presents the results regarding the WTP for the medicine elicited from both groups. It is apparent that mean values are very similar in the two groups.

Table 8. Willingness to pay for the medicine (€).

	Group 1		Group 2
	$WTP(m)^1$	$WTP(m)^1 + WTP(i)^1$	$WTP(m/i)^2$
Mean	24.0	24.63	24.21
Median	20.0	21	25.0
Std Dev	15.322	15.523	10.797
Min.	5.0	5.1	0
Max.	90.0	90.5	50.0

$WTP(m)^1$: WTP for the drug. $WTP(i)^1$: WTP for the brochure (in euros).
$WTP(m/i)^2$: WTP for the whole package including complementary brochure.

Results from the OLS regression analysis performed for explaining the WTP for the medicine is shown in Table 9. It is observed, in a similar way to the case of the WTP for the additional information, the presence of an inverted U-shaped relationship between WTP and age, likely

reflecting the impact of the income variable, which, once again, holds significance in explaining the dependent variable. When the age variable is introduced in the model in the form intervals (see model 4 in the table), a positive and highly statistically significant coefficient ($p < 0.01$) results for the category 35<=Age<50, and a weaker negative effect ($p < 0.1$) for the category "age 65 years or more". This finding suggests a sort of "pensioner" effect, as was already anticipated in section 2.6, since people aged 65 and over in Spain mostly belong to the group of pensioners. According to the Spanish co-payment scheme, pensioners usually pay less for medicines than active population, and consequently they may be more reluctant to declare a high WTP for drugs. Moreover, another finding analogue to that reported from the first regression is the lack of significance in the variables related to the interviewee's prior experience with heart attack and gastrointestinal bleeding, and their numerical skills.

Table 9. Results of the OLS regressions of the WTP for the medicine, WTP(m).

	(1)	(2)	(3)	(4)
Sex	0.723	0.428	1.724	0.202
	(1.814)	(1.721)	(1.755)	(1.815)
Married	-4.719**	-4.819**	-3.769*	-4.532**
	(2.131)	(2.102)	(2.092)	(2.264)
Age	1.672***	1.677***	1.938***	
	(0.302)	(0.299)	(0.273)	
Age2	-0.0200***	-0.0199***	-0.0227***	
	(0.0032)	(0.0032)	(0.00295)	
Heart attack experience	1.800	2.181	1.289	2.435
	(4.153)	(4.082)	(4.174)	(4.288)
Digestive bleeding experience	2.819	2.658	2.717	2.530
	(2.870)	(2.841)	(2.890)	(2.982)
Numerical skills	-2.681	-2.570	-2.459	-2.365
	(2.064)	(1.944)	(2.075)	(2.032)
Questionnaire Type 2	-1.304	-1.594	-0.585	-3.904*
	(2.053)	(1.935)	(2.035)	(2.047)
Income	0.00321**	0.00336**		0.00533***
	(0.0016)	(0.0015)		(0.0015)
Primary studies	-1.257		-0.434	
	(3.640)		(3.642)	
Secondary studies	-1.813		-0.127	
	(4.356)		(4.303)	
Higher education	-0.189		2.459	
	(4.668)		(4.505)	
35<=Age<50				7.634***
				(2.528)
50<=Age>65				0.611
				(2.942)
Age>=65				-5.309*
				(3.127)
Constant	-7.045	-8.571	-10.28	18.00***
	(6.718)	(5.573)	(6.563)	(2.522)
Observations	216	216	216	216
R-squared	0.286	0.284	0.272	0.214

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Regarding the influence of the way that the WTP for the whole medicine package is obtained (as the result of a sequence in Group 1 or through a single valuation in Group 2), the regression analysis upholds the neutrality assumption (the coefficient of the questionnaire variable turns out to be not significant), with the single exception of model 4, for which the coefficient is weakly significant ($p < 0.1$). Parametric and non-parametric tests employed to compare the estimates elicited between the two groups cannot reject the null hypothesis, reinforcing the

impression that apparently there are no 'part-whole' effect in our data (unpaired t-test, p=0.814; WMW test, p=0.863).

Feasibility, test-retest reliability and validity

Feasibility of the questionnaire was assessed by calculating the percentage of completed questionnaires, i.e., questionnaires in which the interviewee answered every single question. In our study, this was the case for 100% of the respondents.

As noted before, a retest was conducted two weeks after the primary interview. To assess reliability over time, the responses of the 20 subjects (10 from each sub-sample) who participated in the retest were compared with their responses in the initial interview. Results are presented in Table 10 indicating that the level of reliability is quite good for the questionnaire administered in Group 1, but not so much in Group 2.

Table 10. Test-retest reliability (correlations).

	Pearson	*Spearman*
Group 1		
WTP for the drug: WTP (m)[1]	0.8630	0.7719
WTP for the information: WTP (i)[1]	0.7219	0.8026
Group 2		
WTP for the drug (inc. Patient information leaflet): WTP (m / if)[2]	0.3375	0.1389
Percentage attributed to information: % WTP (i) [2]	0.3563	0.4679

Regarding (construct) validity, a key theoretical hypothesis for the construct "WTP for the medicine (for the additional information)" is the presence of a positive correlation between the subject's income and their WTP. The results obtained from the regressions presented in Tables 7 and 9 support, in this respect, the theoretical validity of the study. This is evident as the income variable exhibits positive and statistically significant coefficients at the 95% confidence level.

Discussion

The primary objective of this study was to estimate the value of information concerning the benefits and harms of medication based on the best available evidence for effective risk communication. In line with this goal, and according to the main hypothesis formulated in the manuscript, participants in our study valued positively to receive quantitative information regarding effectiveness and side effects of a hypothetical anticoagulant medicine. Our study revealed that the willingness to pay (WTP) for the additional information included in a brochure, complementary to a traditional patient information leaflet, ranged from 60 cents to 1 euro per month. Interestingly, our upper value closely aligns with the average WTP ($1.37) for standardized informational leaflets reported by Dealy et al. (2021) for the United States.

The format used to elicit WTP for the additional information significantly influenced our results, suggesting therefore the existence of elicitation effects. In one of the groups (Group 1) in which the total study sample was divided, WTP for the complementary brochure was elicited separately from WTP for the medication itself. In the other group (Group 2), WTP for the brochure was determined as a percentage of the overall WTP for the entire medication package, including the brochure. Whereas the method used to elicit WTP estimates for both the complementary brochure in Group 1 and the overall WTP for the medicine plus the brochure in Group 2 was a payment card followed by an open-ended question, WTP for the brochure in Group 2 was inferred from the percentage rated onto a 0-100 visual analogue scale (VAS). Our findings clearly indicate that the elicitation format used is not irrelevant, because the mean WTP for the brochure in Group 2 exceeded nearly 60% of the mean WTP elicited in Group 1, even though around one third of participants in Group 2 assigned a zero WTP value to the brochure.

Inspection of the distribution of VAS responses in Group 2 (see Figure A6. in the Appendix 4) clearly shows that, despite the wide variability of the percentages rated (SD = 3.563), there are two values (5 and 0%) in which a considerable volume of the responses concentrates. In this way, 56 (out of 107) participants fitted 5% as the percentage of the WTP for the whole medicine

package that they attributed to the brochure. In contrast, the equivalent percentage (using mean values) that implicitly respondents belonging to Group 1 attribute to the brochure is only 2.6%.

A plausible explanation for this finding comes from the idea that preferences are imprecise, a notion first proposed by MacCrimmon and Smith (1986) to explain preference reversals, then explored further by Butler and Loomes (2007), and that has been also invoked to explain anomalies observed in CV surveys (Dubourg et al., 1994, 1997). The intuition behind this concept is that preferences are (at some extent which varies with the context and task) imprecise by their own, so many respondents in Group 2 would be not sure about the exact percentage to attribute to the brochure, but rather they would have an imprecision interval within which any percentage could represent the relative value to confer to it. As the response scale of the VAS spans from 0 to 100% it is easy that the imprecision interval contained larger values than 2.6%, in such a way that even if responses' WTP for information was identical to that of Group 1, it is conceivable that they perceived a small value as a figure greater than 2.6%, tending indeed to round up their figure to 5%.

Differences between the two respondents' groups in our study also come from the number of zero WTP values obtained in each of them. There was a complete absence of zero responses in Group 1, while a substantial number of null responses (almost 33%) were recorded in Group 2. This result is somewhat surprising, considering that other studies estimating the value of information in non-health contexts (e.g., Latvala and Kola, 2000) have often reported a significant percentage of zero WTP values. It is conceivable, however, that the severity of the health problem addressed in our study (i.e., myocardial infarction) has encouraged participants to pay for getting additional information, thus explaining the unanimity in the responses in Group 1. Notice that, similarly, just one participant in Group 2 was not willing to pay anything for the medicine, including the complementary brochure. So, again, the key of this disparity points towards VAS responses provided in Group 2.

This sharp asymmetry between the WTP for the brochure of the two groups could be attributed to respondents' attention. It is well established in economic theory (Bordalo et al., 2013, 2016) how emphasizing the importance of some attributes can draw consumer attention to them. In the context of our survey there was a marked difference in terms of the salience of the brochure between the two questionnaires that might have led to some respondents of Group 2 to be unwilling to pay anything for it. Although from the onset respondents in the two groups were provided with a complete description of the medicine and the brochure, they were not told in advance that they will be asked to value each in turn. In questionnaire 1, once respondents have stated their WTP for the medicine and before they were asked to state their WTP for the information, they were newly informed of the utility of the brochure to be more awareness of the benefits and risks of the medicine intake. In questionnaire 2, however, immediately after of having state the WTP for the whole medicine package, where the salient attribute was the medicine by itself, and without further elaboration, respondents were inquired about the percentage they would place on the brochure. In this regard, probably respondents' attention was directed to the medicine and not to the brochure, making that many of them underweighted the value of the additional information providing a zero WTP value.

Unfortunately, although participants were encouraged to give thoughtful and reflective answers, no cross-check question was included in the questionnaire to get insight about the actual reasons why some respondents refused to set a positive percentage in the VAS task. This limitation should be overcome in further investigations on this issue.

Apparently, there is no trace of a 'part-whole' bias in our data. The overall WTP for the medicine, which encompasses the brochure, did not significantly differed from the sum of the WTP for the medication alone and the WTP for the additional information. This finding is relevant because there is previous evidence of 'part-whole' effects not only with public goods, but also with private goods (e.g., pizza, desserts), such as the goods valued in our survey are (Bateman et al., 1997; Clark and Friesen, 2008).

We can interpret this result as a consequence of a careful survey design. According to Carson and Mitchell (1995) embedding effects found by Kahneman and Knetsch (1992) were mainly caused by the vague and incomplete description of the goods used in the experiment, and especially because the relationships between them were not specified to all respondents. On the contrary, in our survey the full extent of purchase options was revealed to respondents at the outset. Detailed information on the disease (heart attack) and the side effects (digestive bleeding) of the medicine was described to the respondents, as well as the content of the traditional leaflet and the quantitative information depicted in the brochure. Therefore, at first sight, our attempt to describe accurately the goods under valuation was successful and, accordingly Carson and Mitchell's argument, it could be the reason of the absence of embedding in our survey.

Another perspective more sceptical, however, would be to acknowledge that although in our survey the sum of the part values does not exceed that stated for the whole, it can be due to a sort of composition effect. Strictly speaking a typical part-whole effect occurs when the WTP for the same good when it is valued individually is higher than when the good is valued as part of another more inclusive good (Kahneman and Knetsch, 1992). As saw, mean WTP for the information was significant higher in Group 2 (€1.01) than in Group 1 (€0.63), which is indeed the opposite pattern to that typically reported (Bateman et al., 2006). The implicit mean WTP for the medicine itself in Group 2 (€23.20 = 24.21 – 1.01) is in fact lower than that stated in Group 1 (€24). Therefore, the apparent consistency of the "parts" (medicine and brochure) and the "whole" (the medicine package) hides two effects that offset. This game of communicating vessels results from the application of the VAS mechanism, which, as argued before, may have been able to induce higher valuations through preference imprecision.

The study presented in this paper, while providing valuable insights into the WTP for information about drugs' benefits and side effects, also has several limitations that warrant consideration. Firstly, the sample size utilized in the survey, though sufficient for the specific analysis conducted, may not fully capture the complexity and diversity of preferences within the

broader population. Therefore, it has to be recognized that generalizing the findings to the entire adult population of Spain would be overly ambitious. To address this limitation and enhance the robustness of our knowledge on this issue, future research in this area should aim to employ larger and more diverse samples, striving for greater representativeness of the adult Spanish population. Additionally, studies involving actual patients with varying medical conditions and treatment options could provide valuable insights into the practical implications of information valuation in healthcare decision-making. Exploring the dynamics of patient decision-making processes and factors influencing their WTP for medical information in real clinical contexts would be a worthwhile avenue for further investigation.

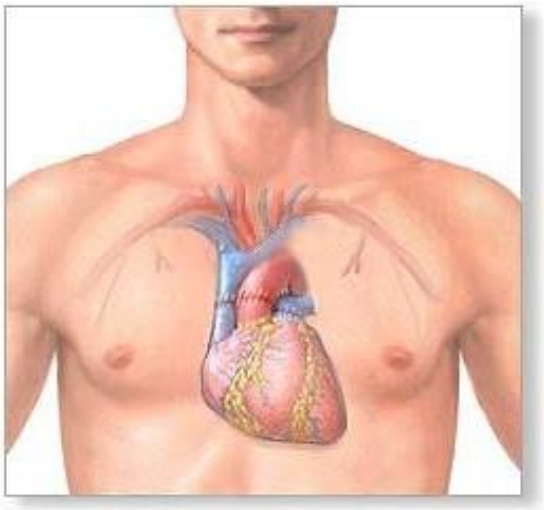
One finding that may at first be surprising, and that could be challenging for the liability of our results, is the high number of subjects lacking numerical skills. We found that just under 36% of the respondents demonstrate sufficient numerical skills to handle small risks, expressed as per 10,000, and convert them into percentages. Notwithstanding, this low percentage is aligned with the findings of previous studies (see Garcia-Retamero et al., 2019 for a review) addressed to test specifically risk literacy. In this vein a paradigmatic study is that conducted by Cokely et al. (2012) who reported results from a test of statistical numeracy and risk literacy (i.e., the Berlin Numeracy Test) administered to highly-educated samples of 15 different countries, Spain amongst them. Just 4% of the Spanish respondents were in the top quartile score, pretty below the average proportion (31%). A similar picture can be extracted from studies testing numeracy in the general population. For example, the first wave of the Programme for the International Assessment of Adult Competencies (PIAAC) study, conducted in 2012, which interviewed over 6,000 Spanish adults (aged 16-65 years), revealed that more than 70% of the respondents had mathematical knowledge at level 2 or lower out of the 5 existing levels (Educainee, 2013).

Despite acknowledged limitations, this paper provides compelling evidence supporting the notion that access to information about the effectiveness and side effects of drugs holds inherent value for the population. Bearing in mind the possibility of extending and improving the study, an interesting implication of the results obtained is that the way of eliciting the value of the

information is not irrelevant. In particular, that asking for such value as a percentage of the value of the drug as a whole may generate biased responses, due to imprecision in preferences and/or the appearance of attention effects. It would be also valuable, certainly, to use alternative methodologies to contingent valuation and engaging real patients as well. In this regard, the use of discrete choice experiments, through which process and outcome attributes are differentiated, arises as a promising approach for future investigations.

Appendix 4

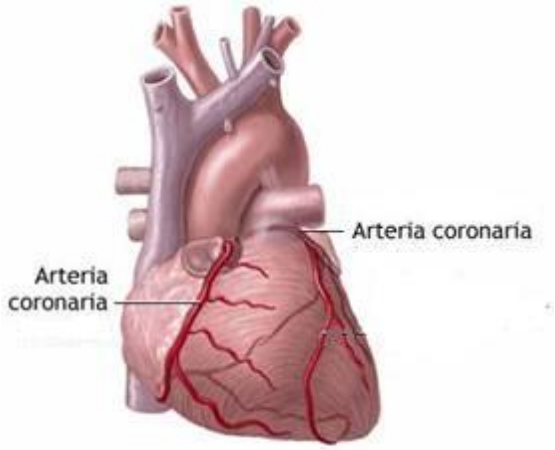
Figure A 3. Card 1: What is a heart attack?



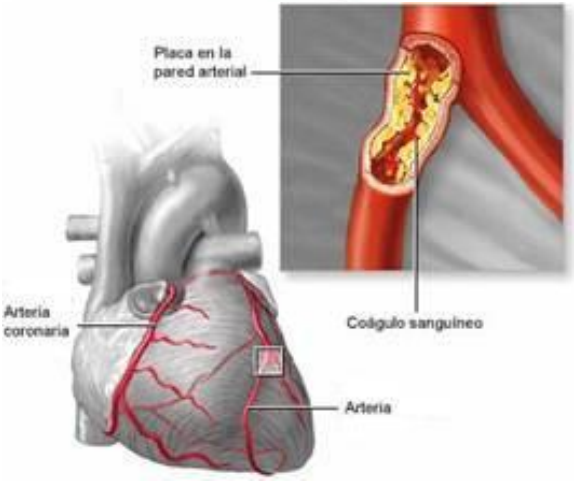
The heart is a composite bag of muscles with veins and arteries that surround it.

It is located between the lungs, on the left side of the thorax, and the muscular mass that forms it is known as the myocardium.

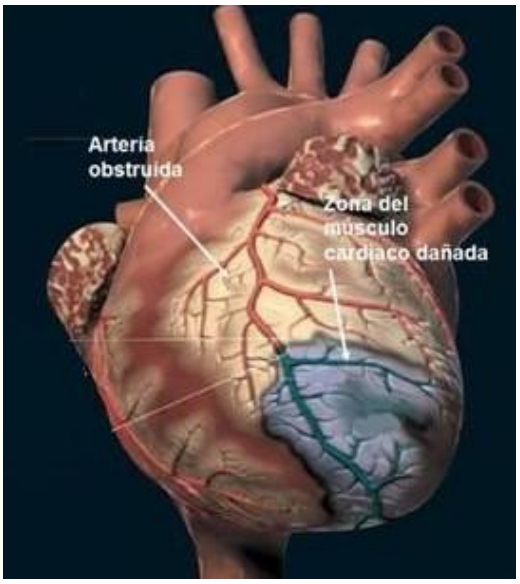
The heart's primary function is to pump blood to all organs of the body to provide them with oxygen.



Like any other organ, the heart also requires oxygen to function properly.
Oxygen reaches the heart through the blood circulating in the coronary arteries, which encircle the heart like a crown.
Proper heart function relies on uninterrupted blood supply from these coronary arteries.



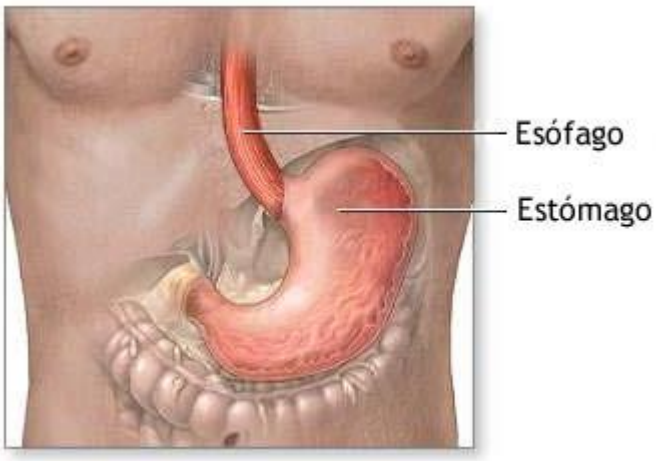
Over time, fat can accumulate in the arteries for various reasons.
If a coronary artery becomes obstructed because of this, the heart does not receive the necessary oxygen.
Permanent oxygen deficiency in the heart muscle can lead to a myocardial infarction, commonly known as a heart attack.
The severity of a heart attack depends on the part of the heart that is damaged.



Smaller infarctions have a higher chance of survival.
Survivors of a heart attack typically require lifelong medication, dietary adjustments, and lifestyle changes.
Unfortunately, between a quarter and a third of people who suffer a heart attack die before receiving medical attention, a condition known as sudden death.

Figure A 4. Card 2:What is a digestive bleed?

The organs of the digestive system, such as the esophagus and stomach, are lined on the inside by a highly sensitive tissue known as mucosa. This digestive mucosa can sustain damage due to inflammatory processes, as seen in conditions like esophagitis and gastritis, or it may result from chronic lesions such as ulcers. Similar to the way external wounds can cause bleeding, injuries to the digestive system can also lead to blood loss, a condition known as hemorrhage.



Gastrointestinal bleeding can be microscopic (only detected in the laboratory) or even massive, in which almost pure blood is expelled.
A small, prolonged bleeding can lead to anemia.
An acute massive hemorrhage can lead to shock due to lack of blood and even death in those who experience it. This occurs in approximately one out of every 10 cases.

Figure A 5. Histograms of the Willingness to pay (WTP) for the information (i.e. the 'brochure') in each of the groups.

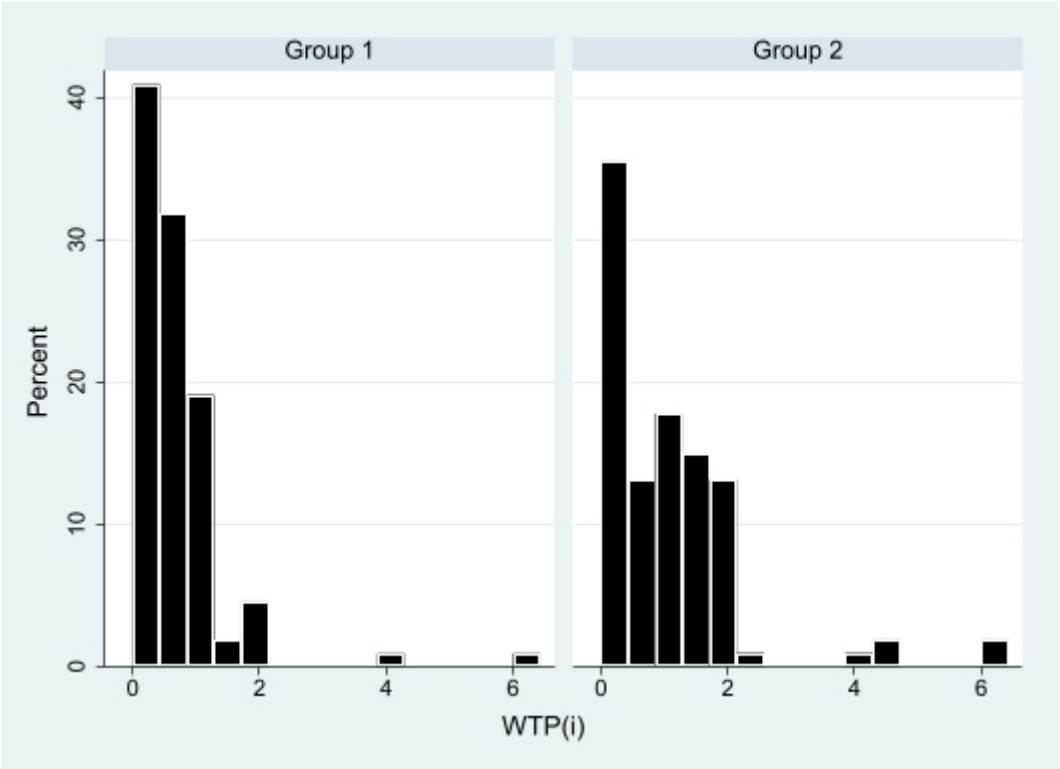


Figure A 6. Histogram of the VAS responses in Group 2: percentages of the WTP for the drug that is attributable to the information contained in the 'brochure'.

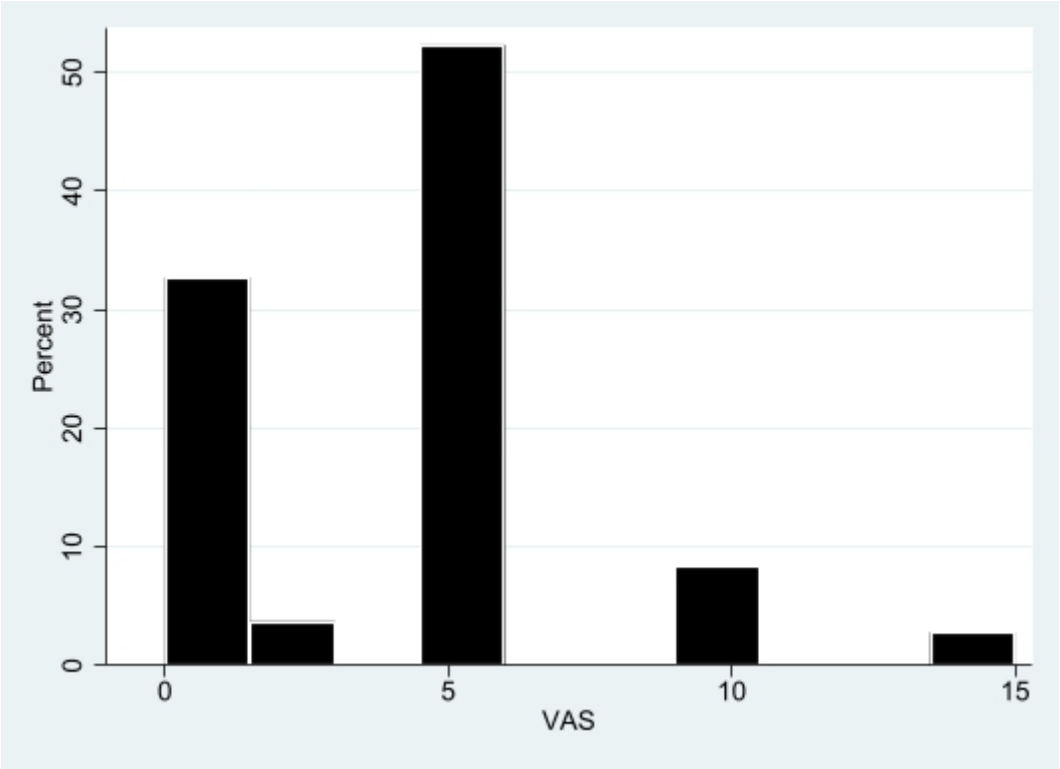


Figure A 7. The "traditional" leaflet.

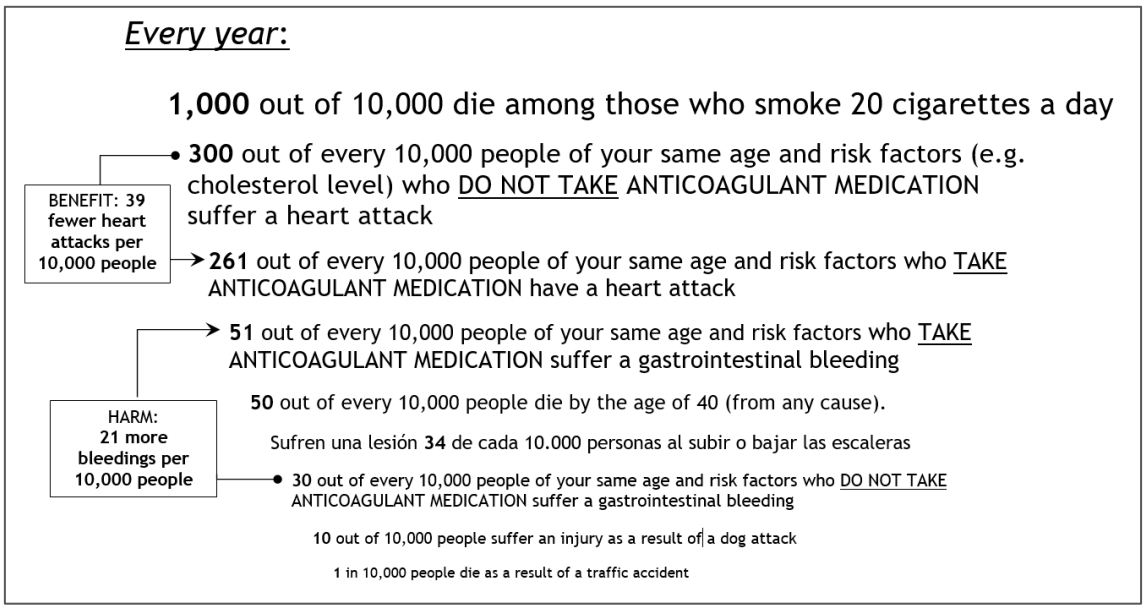
Indications

This medication is indicated for the treatment of ischemic heart disease, specifically coronary artery disease, due to its antiplatelet action. Its anticoagulant effect prevents the formation of thrombi in the arteries, thereby reducing the risk of experiencing a myocardial infarction.

Adverse Effects

Occasionally, individuals may experience gastrointestinal discomfort, including symptoms such as nausea, dyspepsia, and heartburn. Less frequently, adverse effects such as gastric ulcers, duodenal ulcers, and gastric hemorrhage may occur. Gastric discomfort may be more pronounced with higher doses and in patients with pre-existing ulcers or bleeding conditions. Gastric bleeding typically occurs without pain, potentially leading to anemia due to the occult blood loss in feces. The severity of this side effect is associated with the drug's dosage and patient-specific characteristics.

Figure A 8. The complementary brochure. (Additional information on side effects)



4.PAY-FOR-PERFORMANCE IN PRIMARY HEALTHCARE: A REVIEW OF ITS STRENGTHS AND WEAKNESSES

Abstract

Objectives: This study aims to identify the key features an incentive scheme should have to effectively improve primary care professionals' performance in the Spanish National Health Service. **Method:** A systematic search of published studies analyzing the influence of pay-for-performance (P4P) schemes on primary care outcomes was conducted using bibliographic databases (MEDLINE, Scopus, and PubMed) from 1999 to 2019. The pandemic years were set aside, however. Nevertheless, some subsequent studies, considered particularly relevant, were also included in our review. **Results:** The reviewed experiences are largely contingent on the specific context in which they were applied. However, the more successful initiatives share some common features, which are summarized in this article. **Conclusions:** Based on the collected evidence, an effective incentive system for primary care should combine both organizational and financial incentives, provided at both individual and collective levels. The scheme should also be transparent to all stakeholders involved and flexible enough to accommodate the diverse professional situations within the healthcare staff.

Keywords: Pay-for-performance (P4P), primary care performance, healthcare incentives, incentive schemes, organizational incentives, economic incentives.

Introduction

Since the beginning of the present century, public healthcare systems have strived to maintain the quality of care while simultaneously containing expenditure growth. This growth is associated with both demographic factors (e.g., population aging) and demand factors (e.g., patients' increasing aspiration for higher quality care, as they become more informed). Moreover, according to various studies, non-demographic factors affecting the intensity of healthcare use -such as supply-side factors, including technological changes and professionals' behaviour- account for nearly 83 % of the actual rate of increase in health expenditure in Spain according to Abellán-Perpiñán et al.(2013).

It is noteworthy to mention that there is a close relationship between the expenditure resulting from patient demand and that stemming from professionals' behaviour (supply). Thus, in an environment of successive economic crises, such as since 2008, it has been the performance of the healthcare workforce that has enabled the maintenance of medical care without a significant reduction in quality. This dynamic highlights the importance of aligning the goals of healthcare personnel with those of the health system.

Clinical decisions can be influenced by working conditions, particularly by the design of professional compensation, which, in turn, can affect the efficiency of resource use as well as healthcare outcomes. It is apparent that no staff compensation system is perfect. All compensation systems (e.g., fee-for-service, salary, capitation) have their issues. Some systems (capitation and, perhaps, salary) may reduce the level of necessary healthcare provision, while others (fee-for-service) may lead to unnecessary increases in the use of health resources (European Observatory in Health Systems and Policies Series, 2014; Parkin, 2018; Fleetcroft et al., 2012).

In this regard, the American Institute of Medicine (Wolfe A., 2001) stated that, based on the available evidence, the quality of healthcare provided in the US was significantly below best practices, with traditional payment mechanisms being largely responsible for the poor quality of care observed. Consequently, a key recommendation of the report was that payments to

providers should be realigned with the goal of improving quality. This led to a movement in the US in favor of introducing 'value-based purchasing' (VBP) or 'pay-for-performance' (P4P) models in healthcare employee remuneration systems.

The essence of these mechanisms lies in substituting the paradigm of 'paying for doing,' which emphasizes the volume of activity (measured in medical procedures or hours of dedication) as a key element of the remuneration system, with an alternative model where results become central (European Observatory in Health Systems and Policies Series, 2014; Gervás et al., 2008; ANAO, 2007).

The theoretical foundation of P4P programs is based on the so-called doctor-patient agency relationship (Robinson, 2001). According to this theory, the principal (the patient) is unable to adequately assess the cost and effectiveness of available treatments and thus delegates clinical decision-making to their agent (the doctor or the organization for which they work, such as a hospital or primary care center) to ensure the best quality of health services at the lowest possible cost. Since maintaining and improving quality and efficiency require effort from the agent, economic incentives (bonuses) and disincentives (penalties), primarily linked to the achievement of pre-established quality care objectives, are used to reward or penalize this effort, theoretically aligning the agent's interests with those of the principal (European Observatory in Health Systems and Policies Series, 2014).

Australia became the first country to introduce a pay-for-performance system: the Practice Incentives Program (PIP) (Scott et al., 2009; Greene, 2013). It originated in 1999 when a voluntary enrollment period for family doctors was established to align public workers' decisions with the administration's goals in public organizations. Furthermore, P4P systems generally involve linking staff compensation to the achievement of specific objectives.

The Australian experience with P4P was replicated in the US, initially in the private insurance markets (Gemmill, 2008) and later as part of the public Medicare and Medicaid programs. Subsequently, an initiative called the Integrated Health Care Association (IHA), formed in California at the end of 2001 by eight health plans, emerged as a significant program due to its broad scope and became the most important initiative of its kind in the country.

The most ambitious P4P-type incentive program ever launched in Europe is the Quality and Outcomes Framework (QOF), introduced in April 2004 in the United Kingdom. It originated from the General Medical Services (GMS) contract, signed in 2003 between the NHS and the British Medical Association, the professional medical union of the UK. The program aimed to finance general practices or associations where family doctors independently organize the primary care services they provide (Parkin, 2018). This contract, negotiated annually, is funded through three main sources: a risk-adjusted overall capitation endowment, voluntary contributions from family physician organizations for additional services, and additional funding received through the QOF, which raised more than 1.2 million euros in its first three years. The latter is available to groups of family doctors that meet specific quality care criteria established by performance indicators in various areas. Although participation in the QOF is voluntary, more than 95% of these organizations are involved. In July 2018, an assessment of the QOF results was conducted to evaluate potential reforms (NHS Quality-Outcome Framework, 2018). Additionally, a cost-effectiveness study of this incentive system was published in September of the same year (Pandya et al., 2018).

Since its implementation, P4P has accumulated substantial evidence regarding its effectiveness, limitations, and adverse effects. The experience of the British programme in Europe, combined with the absence of similar initiatives in Spain, provides an opportunity to draw valuable lessons for the Spanish healthcare system. In particular, we are interested in their potential application in the primary care setting. Therefore, this article presents a systematic review of various P4P practices undertaken in developed countries over the past decades.

This paper is structured as follows. The next section describes the methodology used to conduct the systematic review of P4P experiences. The following section presents the main findings in line with the British QOF evaluation report. The discussion concludes the manuscript by summarizing key lessons learned from the review.

Methods

First and foremost, to gain a better understanding of the context, a conceptual framework (Figure A9 in the Appendix 5) was designed as a visual representation of the expected relationships between incentives, their effects, and the quality of care. Subsequently, the PICO technique[5] was used (see Figure A10 in the Appendix 5) to construct a well-built research question for this study. This question poses whether there is a P4P model yielding good results in terms of healthcare management and quality of care. Accordingly, a systematic search of publications that analyse the influence of incentives according to the P4P model on primary care outcomes was performed through the bibliographic search engines of Medline, Scopus and Pubmed (see Table 11).

Table 11.Search strategy

1. **SCOPUS:**(TITLE(*"p4p"*) OR TITLE(*"pay for performance"*) OR TITLE (*"incentives"*) AND TITLE (*"primary care"*) OR TITLE(*"practices"*)) AND DOCTYPE(*ar* OR *re*) AND PUBYEAR *>1998*
2. **PUBMED:**("pay for performance"[Title]) OR "incentives"[Title]) AND ("primary care"[Title] OR "practices"[Title]) AND ("1999/01/01"[Date - Completion] : "3000"[Date - Completion])
3. **MEDLINE:**

#7. (#4 OR #5) AND ([article]/lim OR [article in press]/lim OR [review]/lim)

#6. #4 OR #5

#5. #2 OR #3

#4. #1 OR #2

#3. “incentives”/exp/mj OR “incentive”/exp/mj OR “incentive*”:ti

#2. “primary healthcare”/exp/mj OR ((primary NEAR/2(“care” OR “service*”)):ti) OR “unit”:ti OR “units”:ti OR “general practitioner”/exp/mj OR “practitioner”:ti

#1. “pay for performance”/exp OR ((“performance” NEAR/2 “pay*”):ti)

To conduct the systematic review, the following inclusion criteria were applied. First, only articles and reports written in English or Spanish and published from 1999 to 2019 were selected. It was decided to end the systematic search before the onset of the COVID-19 pandemic, as between 2020 and 2022, the implementation of incentive systems was modified in many cases (e.g. the Quality and Outcomes Framework incentive programme for General Practitioners in England) to concentrate resources on responding to the pandemic health

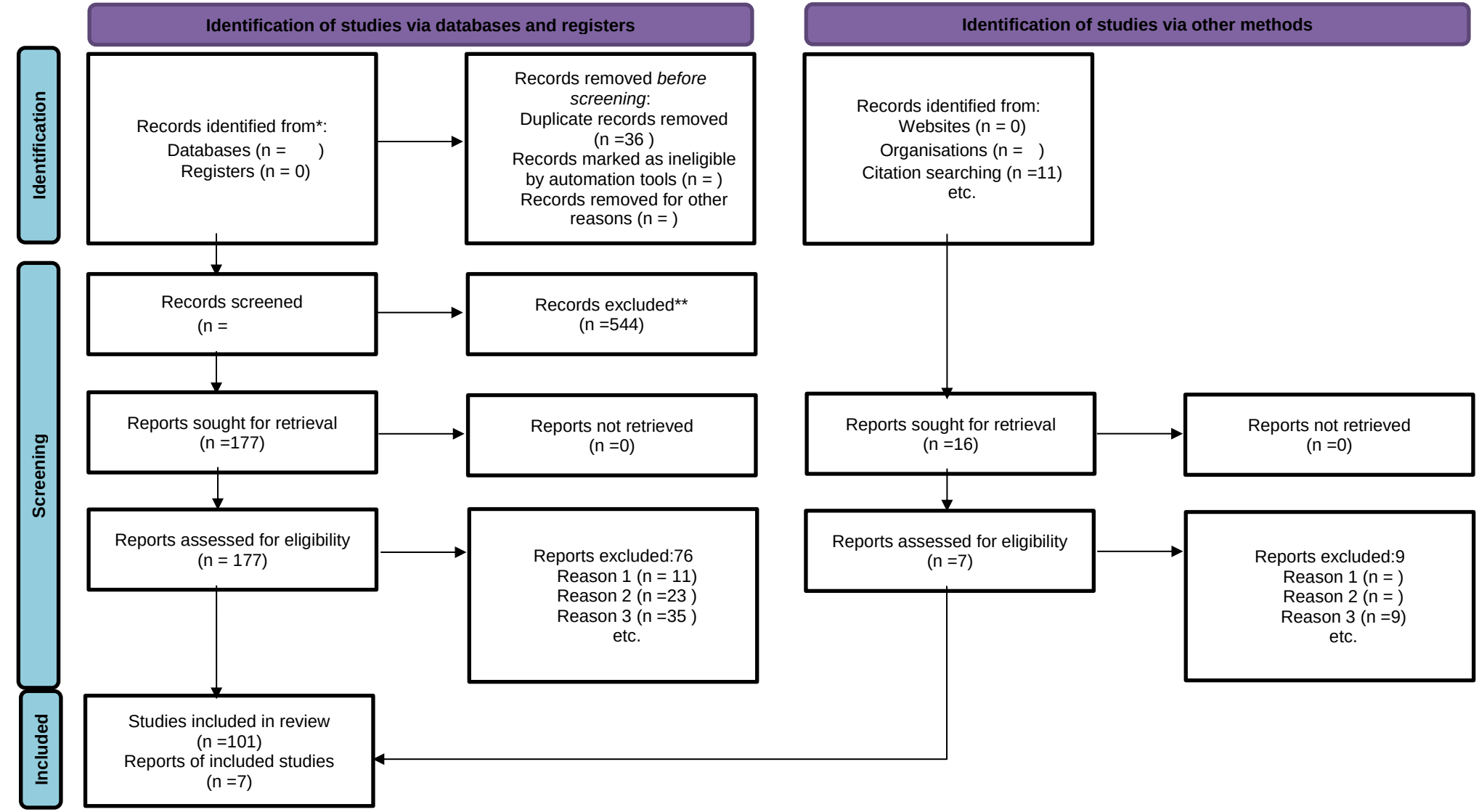
As it is well known PICO (Richardson, 1995) is a mnemonic used to describe the four elements of a well-built foreground question. In this way, P stands for Population/patient, I for Intervention, C for Comparison, and O for Outcomes.

emergency. Second, the selected articles had to report empirical evidence regarding the performance of healthcare professionals' incentive systems. Third, priority was given to experiences in public healthcare systems, although some notable private cases were also considered. Finally, experiences focused on specific work protocols within primary care were excluded.

Lastly, the selection of articles for this review adhered to the PRISMA reporting guidelines for systematic reviews and meta-analyses (see Figure 5)[6]. During the search process, 773 articles were found (757 from a systematic search and 16 from other methods), 36 of which were duplicates. The remaining 737 non-duplicated papers were distributed among three researchers for individual review. Only those articles positively assessed by at least two out of the three researchers were accepted in the first round, resulting in 184 initially selected articles (177 from a systematic search and 7 from other methods). To further narrow down the final number, a second selection round was conducted, leading to the elimination of articles that did not achieve consensus among the researchers. Consequently, a total of 108 articles were chosen (101 from the systematic review and 7 from other methods) and reviewed in depth following the PRISMA guidelines. Figure 5 shows the number of studies added in this way.

[6] The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement originated in 2009 (Moher et al., 2009). It was designed to help authors prepare transparent accounts of their reviews. It was updated to PRISMA 2020, to reflect recent advances in systematic review methodology and terminology (Page et al., 2021).

Figure 5. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



*

Results

Table 12 describes the main features of the selected studies, which indeed share several key ideas that we summarize below.

Firstly, incentives should be simple, straightforward, and aligned with professionals' objectives (Gene-Badía ,2004; Marcotte, 2017;Greene, 2015; Baek, 2013; Kirschner, 2012; Campbell, 2009) Doctors, nurses, and patients felt that the QOF has had a negative effect on continuity of care (Allen ,2014; Gené-Badia ,2007). Furthermore, achieving predetermined activity levels and quality care in fee-for-service, capitation, and pay-for-performance strategies is considered more important than salary remuneration (Ammt et al., 2017; Norman, 2014)

Similarly, it is agreed that small group economic incentives received at the end of the process rarely change health professionals' behaviour (Chung, 2010) making periodic monthly payments more effective. "Most clinicians believe that there should be a mix of clinic and individual-level incentives to maintain collaboration and recognize individual performance" (Greene, 2015). "Offered financial and reputational rewards had statistically significant associations with practice performance" (Allen, 2018; Marcotte, 2017). However, challenges in translating practice guideline recommendations into quality indicators are acknowledged (Robinson, 2001; Institute of Medicine, 2001). "The effect of explicit financial incentives on physician behaviour, it is complicated by a lack of understanding of the incentive structure by the managed care organization and the physician" (Bryan (2001). On the other hand, programs that reward performance in absolute terms may be more effective than those focusing on relative terms (Wright, 2012; Campbell et al., 2009; Johnson et al., 2018), and multiple goals are preferred over single goals (Wright, 2012; Johnson et al., 2018).

In integrated care structures, outcome-based payment programs are more effective when they encompass all providers, combining global budgets and risk-sharing agreements (LaaRocca & Hoholm, 2017; Gené-Badía 2004; Marcotte et al., 2017; Hollander et al., 2015, Lemak,2013). When organized as a traditional P4P system (with a single type of provider and a specific

clinical area), combining individual and group incentives is expected to produce better results than limiting programs to one type of incentive (Greene,2015; Lehtovuori ,2015) .

Table 12. Summary of the selected articles.

Objective of the study	Authors/Year of publication	Location	Main findings
Locate environmental factors that can lead to incremental improvements in contracting systems, according to the information available to improve the comparative setting of objectives and the creation of spaces for good governance and clinical management	J.R. Repullo Labrador y J.M. Freire Campo. SESPAS Report 2024	Spain	Information systems and digital transformation can allow the improvement of management contracts and facilitate comparison competition. Create islands of protected clinical management that recruit innovative professionalism, in whose clinical work systems experiences and consensus are generated to increase internal flexibility.
Taking advantage of the COVID-19 pandemic to overcome inefficient variable pay tied to performance	Gené-Badia J. Aten Primaria/2021	Spain	New quality improvement schemes must also consider improving the patient experience, and a holistic and community-based approach to primary health care.
Studies and reports on health spending	Ministry of Health, Consumer Affairs and Social Welfare	Spain	The main results of the Public Health Expenditure Statistics in different periods of time and its evolution
An analysis of the results obtained from the QOF programme since its implementation in the United Kingdom with a view to its possible reform	England National Health Service(Parkin E.) /2018	England	Evidence suggests that the impact of QOF on health outcomes has been modest at best. There is little evidence to suggest that QOF has had any impact on patient mortality.
Determine the relative performance effects of financial and reputational rewards resulting from a pay-for-performance program	Allen T., Whittaker W., Sutton M., Kontopantelis E./2018	England	Both the offered financial rewards and reputational rewards had small but statistically significant associations with practice performance.
Determine to what extent through public investment and incentives can be improved the implantation for your use of the C-reagent for the treatment of respiratory infection in the trachea	Johnson M.,Cross L., Sandison N.,Stevenson, J.,Moore M./2018	England	Successful implementation can help catalyse aid and exceed operational limits at the local level, but broad implementation will require a change in national policy.
Find out what kind of financial incentives have been employed in mental health centers by the largest mental health service providers in the USA	Stewart.R.E.,MarcusS.C.,HadleyT.R.,Hepburn B.M.,Mandell D.S/2018	USA	Payers are not using those incentives that they perceive as the most effective and employ above all a single strategy for reasons of simplicity and compatibility.
Find out if it is more cost effective to stop or continue the British national health system's primary care performance incentive programme	Pandya A.,DoranT.,ZhuJ.,Walker S.,Arntson E., Ryan A.M./2018	United Kingdom	Continuing the QOF is not cost-effective. To improve health efficiently, the UK should redesign QOF or pursue alternative interventions.
Examine how the transition from a pay-per-service model to a mixed capitative, per-service model occurs in primary care teams	Zhang X.,Sweetman A./2018	Canada	Doctors reduce the provision of capitative payment codes with no change in FFS(Feeforservice) services. All other doctors reduce both capitative and FFS services.
To assess the effect of incentive payments on chronic disease management and health care use at Columbia	Lavergne M.R.,Perterson S.,GarrisonS.,Hurley J., Cheng L., McGrail K./2018	Canada	This large-scale incentive program, does not produce changes in patient management, reduced hospitalizations or changes in spending on patients with diabetes and COPD are observed.
The impact of financial incentives on pioneers and supporters among U.S. hospitals. Observational study	Bonfrer I, Figueroa J F, Zheng J, Orav E J, Jha A K./ 2018	USA	Hospitals that pioneered the use of financial incentives have not reduced mortality much more than those that started using them much later, due to the low value of the incentive and the complex structure of the indicators chosen for monitoring..
It analyzes the increase in overtime performed by doctors and how they are affected by personal and family circumstances as well as the incentive structure.	Ammt M.,Fortier G./2017	Several European and North American countries	Increasing incentives appears relatively ineffective to the now of encouraging more overtime and can be detrimental if the incentives are not well placed. Welfare systems contribute to explain variations in P4P experiences.
Examine the effects of new measures introduced to improve coordination: a tool based on the ICTT protocol and an incentive programme	LaaRocca A.,Hoholm T./2017	Norway	The inter-organisational nature of coordination in the health sector makes it crucial for managers and policy-makers to devise incentives and instruments that work across different organisations rather than just in one organisation.

Find out how financial incentives can change the behaviour of primary care physicians	Marcotte L.,Hodlofski -A.,Bond A.,Patel P.,Sacks L.,Navathe A.S./2017	USA	Non-financial incentives used in conjunction with financial incentives can reinforce and redirect the shortcomings of each type of incentive.
It is about demonstrating that the implementation of the new incentive program has caused changes in practices, generating competition to see who are the best doctors	Fichera E.,Pezzino M./2017	England	Following the introduction of the program, efficient doctors are likely to become companions and mobility among physicians has increased.
We examined the impact of the incentive program for the treatment of complex patients, on access and continuity of care in primary care, admission ratios en hospitales y costes	Lavergne M.R.,PertersonS.,GarrisonS.,Hurley J., Cheng L., McGrail K./2016	Canada	This incentive program does not appear to have improved access and continuity of care, or reduced excessive use of resources elsewhere in the health service.
Better understand the consequences of being exempted from treatment due to the incentive structure and its relationship to patient type and mortality	Kontopanelis E.,Springate D.A.,Ashcroft D.M., Valderas J.M., Van Der Veer S.N. Reeves D.,Guthrie B., Doran T./2016	United Kingdom	Older patients with multimorbidity are more likely to be exempted from treatment within the incentive structure.
To study the extent to which the incentive structure of the British QOF is associated with a reduction in population mortality.	Ryan A.M.,Krinsky S., Kontopanelis E., Doran T./2016	United Kingdom	Although small reductions in mortality were observed in groups composed of incentive target patients, QOF was not associated with significant changes in mortality.
Assess the impact of pay-for-performance on healthcare staff behaviour and patient outcomes	Lin Y.,YinS.,HuangJ.,Du L./2016	United Kingdom	Decision-makers should consider establishing minimum conditions of quality and clinical practice before establishing new policies.
Determine whether German healthcare workers would participate in a pay-for-performance system	Krauth C.,Liersch S.,Jensen S. , Amelung V.E./2016	Germany	. The crucial factor for (not) accepting P4P might be the sense of (un)fairness of P4P..
Describe the development of methods used to evaluate the cost-effectiveness of pay-for-performance indicators and how this has contributed to the emergence of new indicators	Quereshi N.,WengS.,Hex N./2016	England	The current methods used to assess the feasibility of new economic indicators could change their national approach to a more local one while maintaining central supervision.
Gain understanding of the views of primary care patients in the French system on pay-for-performance	Olivier Saint-Lary, Claire Leroux, Cécile Dubourdieu, Cécile Fournier and Irène François-Purssell/2015	France	Since the implementation of pay-for-performance, patients have not noticed any changes in clinical care.
Determine whether preventive care activities in general medicine are sustained when financial incentives or information from external audits disappear	Hocking J.S., Temple-Smith M.,van Daniel M.,LawM.,Guy R.,BulfoneL.,WoodA.,LowN.,DonovanB.,FairleyC.K.,Gunn J./2016	Australia and New Zealand	The results of this trial will have implications to support prevention measures for health.
Compare and contrast the mechanisms of financing incentives in general practices in a rural environment	Neil A.L., Nelson M.,Palmer A.J./2016	Australia	The rapid introduction of the new incentive model has led to the emergence of inconsistencies and exacerbated inherent biases, particularly inequality towards rural providers.
Assess the impact on physicians' prescribing behaviours of an outpatient prescribing incentive program on South Korea's primary care	Park S.,Han E./2016	South Korea	The outpatient incentive programme has operated as planned in certain clinical subgroups for the target drugs. The degree of prescription in the centres has been reduced
To assess the effect of group incentives in primary care on outcomes across the different disciplines of the clinical team	Lehtovuori T., Kauppila T.,Kallio J.,Raina M.,Suominen L.,Heikkinen A.M./2015	Finland	Group incentives can provide a method for altering clinical practices in primary care. However, the sustainability of these interventions may be reduced after these incentives cease.
Assess the impact on annual health costs and hospital utilization patterns caused by the incentive system at Columbia	Hollander M.J., Kadlec H./2015	Canada	Although the current literature shows mixed results on pay-for-performance, it was shown that the system used at Columbia in primary care could reduce healthcare and hospital utilization costs.
Demonstrate whether primary care personnel are more likely to participate in performance incentive programs than other specialties	Russel T.,Petterson S., Klink K., Bazemore A./2015	USA	Primary care personnel are more likely to participate in performance incentives after taking into account the baseline income of each of the specialties.
Evaluate the results of different incentive programs in primary care	Guilliam S./2015	United Kingdom	Mixed results are obtained derived both from the existing organizational system in each country and the historical context

To demonstrate that the mortality rate in England between 2007-2012 was not associated with the quality indicators included in the P4P program, QOF, implemented and England	Fleetcroft, Robert/2015	England	The mortality ratio was not associated with the quality indicators established by the QOF program
Examine the adoption of financial incentives in practice and identify which type of practitioners are most likely to participate in these programs	Kecmanovic,Milica; P.Hall, J. /2015	Australia	It is therefore crucial to consider such costs(administrative burden to claim incentives) in relation to the size of the payment when designing incentives, and not worth the effort.
Assess the impact of the QOF program on care outcomes and quality	Guill, Paramjit/2015	United Kingdom	QOF indicators have been for the most part successful while focusing on the processes and structures needed for quality. However, many of the outcomes rely excessively on process measures.
To examine how six months of interventions in selected practices impacted best practices and outcomes on minority patients receiving primary care in medical practices subject to a P4P programme	Johnson R.M.,Johnson T.,Zimmerman S.D., Marsh G.M., García-Dominic O./2015	USA	The interventions with the greatest impact were those related to face-to-face educational conversations, clinical documentation of the patient on paper rather than those aimed at sustaining adherence to treatment.
Examine the driving points of mechanisms to report on a correct, transparent and auditable development of an overtime incentive programme in Tasmania	Neil A.L., Nelson M.R.,RichardsonT.,Mann-Leonard M.,Palmer A.J./2015	Australia(Tasmania)	The incentive program does not preferentially support practices that provide overtime care and arguably lead to perverse incentives.
Discuss the structure and components of the pay-for-performance program in the quality of primary care care	Allen T., Whittaker W., Masón T. /2014	United Kingdom	If P4P is to last in the long run, the question about the cost-effectiveness of the system must be answered in a resounding way.
Investigate whether general practitioners who are part of the pay-for-performance program also met those indicators not included in the incentive program	Fernández Urrusuno R.,Pérez Pérez P., Montero BalosaM.C.,Márquez Calzada C.,Pascual de la Pisa B./2014	Spain	The degree of compliance demonstrated by general practitioners with indicators outside the incentive program was no different than with those who were within it.
To investigate whether the experience of health workers in primary care under the local incentive system (PCT) is different from that obtained under the new QOF incentive programme in primary care	Hackett J.,GlidewellL.,West, R.,CarderP.,DoranT.,Foy R./2014	United Kingdom	The nature of pay-for-performance was not reduced by the local incentive structure.
Exploring the effects of the British P4P system on primary care	Norman A.H.,Russell A.J., Macnaughton J/2014	United Kingdom	A bureaucratic-scientific model that has profound effects on primary care in England has been gradually reinforced.
Paying for performance in Healthcare: Implications for Health System Performance and Accountability	European Observatory in Health Systems and Policies Series/2014)	Europe/ OECD publishing	Different results according with the countries and context
Examinar la participación de grupos pequeños y medianos de médicos en un programa de P4P e informar sobre el programa y las practicas incluidas dentro del mismo	Casalino,Lawrence P./2013	USA	Small- to medium-sized practices appear to respond to PR and financial incentives by greater use of CMPs.
Describe and compare the recently existing indicators for primary care and their related funding in European countries	Kolozsvari L.R.,Orozco-Beltran D., Rurik I./2014	European countries(European General Practice Research Network)	The implementation of these programs should be critically evaluated with continuous monitoring at the national and regional levels.
Study and analysis of how the P4P incentive system influences the performance of the health system	Cheryl Cashin, Y-Ling Chi, Peter Smith MB and ST -European Observatory in Health Systems and Policies Series (2014)	12 OECD countries (including Estonia, France, Germany, Turkey and the United Kingdom)	The results are diverse, depending on the health system studied, since the starting or basic conditions are different in each case.
To examine the participation of small- and medium-sized physician practices in pay-for-performance and public reporting programs and the characteristics of the participating practices	Hearld LR, Alexander JA, Shi Y, Casalino LP (2013)	USA	The study suggests that some structural features are associated with participation and may provide leverage points for fostering participation.

Study why most French primary care physicians chose not to join the voluntary pay-for-performance program	Saint-Lary O.,Bernard E.,Sicsic J.,Plu I., Francois-Purssell I., Franc C./2013	France	The perception of ethical risks implicit in pay-for-performance may have hindered its success.
Describe the quality indicators evaluated for the improvement of this in Europe through incentive programs based on payment for performance in primary care	Kolozsvari L.R.,Rurik,I-/2013	United Kingdom, Hungary and other European countries	It is observed that the range in which incentives move usually goes between 1%-25%. More data is required in primary care in order to improve the indicator system.
Explore the practices of primary care physicians with pay-for-performance in the Netherlands	Kirschner K.,Braspenning J.,Jacobs J.E.A., Grol R./2013	Netherlands	.It is considered that P4P stimulated quality improvement but that it could have unforeseen consequences. It was suggested to link part of the bonus to innovation
Gain a greater long-term perspective of P4P implementation	Lester H.,Matharu T.,Mohamed M.A, Lester D.,Foskett-Tharby R./2013	England	Calibrating the appropriate level of clinical autonomy is critical if pay for performance schemes are to have maximal impact on patient care.
Assess changes in performance after introducing a P4P program	Kirschner K.,BraspenningJ.,Akkermans R.P.,Jacobs J.E.A., Grol R./2013	Netherlands	A participatory P4P program might stimulate quality improvement in clinical care and improve patient experiences with GP´s functioning and the organization of care.
To examine whether aligning economic incentives with daily health practice continues to provide high-quality care	Baek J.D., Xirasagar S.,Stoskopf C.H.,Seidman R.L./2013	USA	Financial incentives aligned with clinical productivity/benefit do not impede high-quality care. Incentives associated with indicators can help transform and improve the quality of care.
Assessing the impact of rural provider services mix on the Primary Care Incentive Payment Program	Shane D.,MacKinney A.C.,Ullrich F.,Mueller K.J., Weigel P./2013	USA	P4P can have a significant effect on the adequacy of aggregate indicators. Evidence indicates that a tunnel vision effect is possible when public authorities monitor specific sets of indicators.
To examine the effect of public reporting and financial incentives linked to performance quality in small and medium-sized medical groups	Alexander J.A., Maeng D.,CasalinoL.P.,Rittenhouse D./2013	USA	Small and medium-sized practices appear to respond to public reporting and financial incentives. Future research is required to figure out the right mix of incentives and reports.
Examine the impact of the P4Pen Australia program	Greene, Jessica/2013	Australia	The centres reported that this incentive had not altered their behaviour mainly because the amount of the payment was modest and it was difficult to record the data.
Effects of pay for performance in health care: A systematic review of systematic reviews	Frank Eijkenaar, Martin Emmert, Manfred Scheppach, Oliver Schöffski /2013	Global	Twenty-two reviews contain evidence on a wide variety of effects. Findings suggest that P4P can potentially be (cost-)effective, but the evidence is not convincing
To assess the effect of low P4P remuneration on patients' health	Houle S.K.D., McAlister F.A., Jackevicius C.A,Chuck A.W., Tuyuki R.T/2012	Canada	The effect of P4P on patients' health remains largely unknown or uncertain. Its implementation must be accompanied by robust evaluation plans.
To update the existing evidence on P4P in the quality of primary care and present the usual problems of primary care	Wright M./2012	Australia	More research is needed to assess the effect of these programs on health and healthcare outcomes
Review existing evidence on the impact of P4P programmes on primary care outcomes and quality	Gilliam S.J., Siriwardena A.N, Steel N./2012	United Kingdom	Modest improvements in health outcomes were observed, in terms of cost effects, professional behaviour and patient experience remain uncertain
Verify that indicators with greater health benefits received larger incentives	Fleetcroft R.,Steel N.,Cookson R.,Walker S., Howe A./2012	United Kingdom	No associations were found between incentive size per outcome in an indicator and expected improvement in health. The indicator should be aligned to the expected gain in health.
Assess the funding needed to support a primary care incentive program taking into account the risks of each type of patient	Ash A.S., Ellis R.P./2012	USA	Existing data can support risk-adjusted payment calculations and appropriate performance assessments needed to power transformations in primary care
Develop a framework in order to study the mechanisms of implementation of an incentive program, the means by which medical practices and physicians translate the	Cohen G.R.,Erb N.,Lemak C.H./2012	USA	Unclear points were identified in previous research regarding how incentives influenced medical practices.

objectives of the incentive program into a specific official framework			
To study the effect of two major changes in subsidies for general medical services in Australia.	Connelly L.B.,Butler J.R.G./2012	Australia	There is evidence that in the short term the supply curve for general medical services is curved backwards.
Describe the growing ethical tensions in France	Saint-Lary O.,Plu I.,Naiditch M/2012	France	No associations were found between incentive size per outcome in an indicator and expected improvement in health. This disconnect between incentive and gain in health supports activities that are marginally effective.
Evaluate the results obtained in a P4P program developed by its own users	Kirschner K., Braspenning J., Jacobs J.E, Grol R./2012	Netherlands	The resulting program is aligned with the target users' views on priority and usefulness. This can increase the degree of commitment to the program.
Gaining primary care patients' perspective on P4P in Britain	Hannon K,L., Lester H.E., Campbell S.M./2012	United Kingdom	This study adds the patient's point of view to the debate about the consequences of P4P in health care, the concern of users revolves around how this program is being evaluated.
To examine the effect of changes in the method and level of disbursement on the quality of primary detention.	Scott A.,SiveyP.,AitOuakrim D.,Willenberg L.,Naccarella L., Furler J., Young D./2011	International(Several data bases)	There is insufficient evidence to support or not support the use of financial incentives to improve the quality of primary health care.
Describe the quality of care for patients with chronic diseases among older people in care facilities	Shah S.M.,Carey I.M., Harris T.,DeWildeS.,Cook D.G./ 2011	United Kingdom	There is scope to improve the management of medical centers for chronic diseases, but some P4P systems do not disadvantage patients compared to those who live in their community
Discuss the results of existing literature reviews on P4P programs	Capizzi, Silvio/ 2011	International-Sweden,Norway, UK, Netherlands New Zealand,USA.	It has been shown that the quality of care and is higher in countries where performance is monitored and there are control systems associated with economic incentives.
An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviors and patient outcomes	Gerd Flodgren, Martin P Eccles, Sasha Sepperd et al. (2011)	Global(Cochrane Database)	Financial incentives may be effective in changing healthcare professional practice
Examine experiences of other countries in the implementation of P4P and ask how to transfer these experiences to Australia	Campbell S.M., Scott A., Parker R.M., Nacarella L., Furler J.S., Young D., Sivey P.M./2010	Australia	Several important lessons must be learned from other countries if the system is to be implemented in Australia.P4P should be used as part of a wider strategy for quality improvement..
To estimate the potential reduction in population mortality after the implementation of P4P in England	Fleetcroft R.,Parekh-Bhurke S., Howe A., Cookson R.,Swift L., Steel N./2010	United Kingdom	P4P has brought a gain in health, but potential improvements in health were limited by performance targets for full pay that were in a smaller range than the typical performance measure.
To assess the effect of a P4P programme on the quality of preventive care in a network of community health centres	GavaganT.F.,DuH.,Saber B.G., Adams G.J.,Graham D.M., McGray R., Goodrick G.K/2010	USA	There was no evidence for a clinically significant effect of financial incentives on performance of preventive care in these community health centers.
Generate a draft of the relationship between P4P programs and quality improvement	Peckham S., Wallace A./2010	United Kingdom	P4P programs need to take more into account broader concepts of quality. Future incentives should be balanced against sanctions for poor practice.
Quantify young doctors' preferences for different attributes relevant to practice in Germany	Günther O.H.,Kürstein B.,Riedel-Heller S.G., König H-H./2010	Germany	Results indicated that a change in income led to the largest utility change compared with changes in other attributes
To assess the effect of a specific P4P programme on measures of quality of care in large group practices	Chung S., Palaniappan L.P., Trujillo L.M., Rubin H.R.,Luft H.S/2010	USA	Physicians' responses to a P4P program with a small maximum bonus do not differ by frequency of bonus payment
Check how a P4P program affects the health care of racial and ethnic minorities below socioeconomic status	Friedberg M.W., Safran D.G., Coltin K., Dresser M.,Schneider E.C/2010	USA	P4P programs should monitor and direct the potential impact of pay-for-performance toward healthcare disparities
Assess the robustness of patients' responses to a national survey of their patient experience as a basis for providing financial incentives to doctors	Roland M., Elliot M.,Lyratzopoulos G.,Barbiere J., Parker R.A., Smith P.,Bower P.,Campbell J./2009	United Kingdom	There is little evidence to support the concern of some general practitioners that low response rates and selective non-response bias have led to systematic unfairness in payments attached to questionnaire scores .
To assess the effects of the implementation of the P4P	Campbell S.M., Reeves D.,Kontopantelis	United Kingdom	Financial incentives are most likely to be an effective means of influencing

program in primary care	E.,Sibbald B., Roland M./2009		professional behavior when performance targets and rewards are aligned to the values of the staff being rewarded
Evaluate the changes and describe the experience of a P4P linked to quality objectives in two health centers in Spain	Benavent J., Juan C.,Clos J., Sequeira E.,Gimferrer N., Vilaseca J./2009	Spain	Several area needing improving were detected: process-management training for healthcare professionals, process orientation to the patients, the need to develop a communication plan, the selection of process and outcomes indicators, appropriate use of information systems and the time spent in implementation of the quality-improvement programme.
Describe and analyze the impact of an incentive model linking disbursement to adherence to medication and therapy committee guidelines, as well as a self-reflection on prescribing pattern in a report	Wettermark B.,Pehrsson A.,Juhasz-HaverinenM.,Veg A.,Edlert M.,Törnwall-Bergendahl G.,Almkvist H.,Godman B.,Granath F., Bergman U./2009	Sweden	Although no causal effect can be attributed without a control group, we have shown the feasibility of a model linking payment to DTC adherence.
Assess changes in the size and composition of primary care teams, as well as their workload with the introduction of the P4P incentive program	Gemmell I.,Campbell S.,Hann M.,Sibbald B./2009	England	The number of workers increased, mainly that of nurses compared to that of doctors. Nurses absorbed most of the workload caused by the new program while doctors focused on chronic care and prevention.
Study the potential unintended consequences of P4P programs in England and California	Mcdonald R.,Roland M./2009	England and California	It is suggested that the unintended consequences of P4P relate to the way these programs are designed and implemented.
Assess the impact and the new P4P program and new incentives	Grant S., Huby G.,Walkins F.,Checkland K., Mcdonald R.,Davies H.,Guthrie B./2009	United Kingdom	The health person seems to have adopted the new conditions while maintaining the quality of care.
The effects of an incentive program on quality of care in diabetes management	Scott A,Schurer S, Jensen PH, Sivey P./2009	Australia	The study finds that the incentive program increased the probability of an HbA1c test being ordered by 20 percentage points
Examine the relationship between socioeconomic inequalities and the quality of care performed	Doran T.,Fullwood C.,Kontopantelis E.,Reeves D./2008	England	The results support that economic incentives have the potential to help reduce inequalities in the quality of care in the most underserved areas.
Investigate the mechanisms and perceptions of control that follow the implementation of a new P4P contract	McDonald R.,Harrison S.,Checkland K./2008	United Kingdom	Increased surveillance and feedback mechanisms associated with new pay-for-performance schemes have the potential to constrain and shape clinical practice.
Quantify the impact of the latest GMS contract model, in Scotland, based on the QOF	McBride-Stewart S.P.,Elton R., Walley T./2008	Scotland	The prescription of relevant drugs increased with the introduction of MSG, the increase continued in the first two years of the new contract but at a significant lower level.
Pay for performance	Gérvas, J., Pérez Fernández, M/ 2008	Spain	Incentives must be selected that increase the health of patients and the population.
Pay-for-Performance in the US: What lessons for Europe?	Gemmill M./ 2008	USA	Based in the experience in the USA, new considerations are then discussed in light of European health policy, and provide insights for European policy makers on the adoption of P4P programmes.
Investigate the degree of improvement in the quality of care after the implementation of a P4P system	Campbell S.,Reeves D.,KontopantelisE.,Middleton E., Sibbald B., Roland M./2007	England	There has been an improvement in the quality of care for conditions linked to objectives in the short term. However, continuity of care was reduced after the introduction of the scheme.
Explore the impact of financial incentives on healthcare quality, clinical autonomy and internal staff motivation	McDonald R.,Harrison S.,Checkland K., Campbell S.M., Roland M./2007	United Kingdom	The implementation of these programs does not seem to have damaged the degree of motivation of the staff studied, although more concern was expressed by nurses.
To assess the initial effects of a behavioural-independent incentive on the prescribing level of primary care physicians	Martens J.D., Werkhoven M.J., Severens J.L.,Winkens R.A.G./2007	Netherlands	Behaviour independent financial incentives can be a help in changing prescription behaviour of GPs, but effects are small-scale and temporary.
To examine the relationship between changes in quality recorded for four chronic afflictions in the periods 2003-2005 and incentive payments	Steel N., Maisey S., Clark A.,FleetcroftR.,Howe A./2007	United Kingdom	The introduction of incentives was apparently associated with an improvement in quality for the incentivized conditions. For non-incentivised conditions, quality did not appear to improve.

Assess whether the implementation of incentive programmes has had an impact on the quality of professional life of health workers and on user satisfaction	Gené-Badia J.,Escaramis-Babiano G.,Sans-Corrales M.,Sampietro-Colm L.,Aquado-MenguyF.,Cabezas-Peña C.,Puelles P.G./2007	Spain	Incentives linked to annual quality objectives can increase the sense of burden of health personnel and can have a negative impact on user satisfaction. Incentives in long-term professional development are related to an increase in the perception of the support professional by management.
February Report (2007)	Australian National Audit Office (ANAO)	Australia	The ANAO recommends that agencies incorporate into business or budget planning processes, a periodic review of outcomes and outputs
Explore the relationship between financial incentives and likely health gains	Fleetcroft R.,Cookson R./2006	United Kingdom	The disbursements in the P4P contract do not reflect the health gains of the population.
To determine to what extent the practices that had obtained high scores, was thanks to classifying patients as ineligible for the quality indicators or exception report.	Tim Doran, M.P.H., Catherine Fullwood, Ph.D., Hugh Gravelle, Ph.D., David Reeves, Ph.D., EvangelosKontopantelis, Ph.D., Urara Hiroeh, Ph.D., and Martin Roland, D.M(2006)	United Kingdom	A small number of centers appear to have achieved high scores by excluding large numbers of patients through the exception report.
Examine the role in practice that executives play in implementing P4P programs and how their perspectives and decisions can influence the success of these programs.	BokhourB.G.,Burgess Jr. J.F.,Hook J.M., White B., Berlowitz D., Guldin M.R., MeterkoM.,Young G.J.(2006)	USA	5 key points were identified: Incentives on quality are better than on use or use, are an additional reward, are agents of change and the ways in which quality is measured are problematic.
Describe the initial impact of the new set of P4P incentives	Roland M.,Campbell S.,Bailey N.,Whalley D., Sibbald B./2006	United Kingdom	GPs are employing more staff, especially nurses and data entry clerks, and computerizing their clinical records so it can exist unintended consequences.
Propose a new incentive model for primary care physicians	Moreno V.M.,Cermeño P.C., Gadea J.E., Vicente V.B./2006	Spain	An open and dynamic incentive model must be created that takes into account the variability of working conditions. 7 modules will be evaluated (quality of care, rational use of medication, generics, pharmacy, efficiency associated with clinical practice, continuity of training and teaching, research
Find out the relationship between financial incentives, the selection of prescription indicators and the success of staying within budget	Ashworth M.,Lea R.,Gray H.,Rowlands G.,Gravelle H.,Majeed A./2003	England	Large prescription incentive programmes may have helped to control costs, but their effect on prescribing quality is uncertain.
Variable payment linked to quality of care	Gené-Badía J, G. de P. P. (2004).	Spain	Economic incentives must be framed in a global reform project and must be applied with a participatory, simple, transparent and objective methodology, using indicators that reflect relevant aspects of care. The results of user satisfaction surveys have not changed significantly compared to records from previous years. In the future, it is necessary to consider the value that this indicator may have to effectively measure the quality of professional care.
Find out the level of influence of primary care professionals when developing an incentive program and whether a position of responsibility in the organization gives influence over their colleagues within the same field. citizen.	Mahmood K./2003	England	The attitudes of health personnel working with the citizen with respect to their colleagues in management reinforce notions of a new division in general medicine.
Report: Crossing the Quality Chasm: A New Health Care System for the 21st Century. Policy, Politics, & Nursing Practice	Institute of Medicine (US) Committee on Quality of Health Care in America/ 2001	USA	Payments to providers should be realigned with the goal of improving quality
Theory and practice in the design of physician payment incentives	Robinson J. / 2001	USA	The complexity of healthcare services makes it difficult to establish a fair framework for the professionals an the patients
The effect of explicit financial incentives on physician behavior	Bryan S. Armor (2001)	USA	The effect of explicit financial incentives on physician behavior is complicated by a lack of understanding of the incentive structure by the managed care organization and the physician

Identify all financial incentives that have been proposed, described or used regardless of their initial objective and, where possible, assess the results of these incentives in the costs, processes or results of care	Chaix-Couturier C., Durand-ZaleskiI.,Jolly D.,Durieux P./2000	International(Several data bases)	It may be effective to use incentives in combination depending on the target set for a given health care programme
Analyze the impact on costs and care results of a group incentive program for primary care in Michigan	Lemak C.H., Nahra T.A., Cohen G.R., Erb N.D.,PaustianM.L.,Share D., Hirth R.A./2013	USA	An economic incentive together with an audit and another tool to obtain information on the results were effective in modifying the behaviour of the staff, improving the quality of care
Evaluate the impact on healthcare outcomes of the existing economic incentive at the Fairview healthcare provider	Greene J., KurtzmanE.T.,Hibbard J.H., Overton V./2015	USA	There is evidence of the potential effectiveness of aligning payments with costs and quality in performance.
To determine whether the withdrawal of the Quality and Ooutcomes Framwork(QOF) scheme in primary care in Sccotland in 2016, had an impact on selected recorded quality of care, compared with England where the scheme continued	Daniel R Morales, Mark Minchin, Evangelos Kontopantelis, Martin Roland, Matt Sutton, Bruce Guthrie (2023)	Scotland	The abolition of financial incentives in Scotland was associated with reductions in recorded quality of care for most performance indicators.

Next, we classify the selected studies according to five criteria: the four used by the NHS Quality-Outcome Framework Report (2018) to assess the pros and cons of P4P schemes in England, plus an additional criterion, cost-effectiveness, based on Pandya et al.'s (2018) paper. The five criteria are as follows: impact on patients, organization, unintended consequences, withdrawal of indicators, and cost-effectiveness

Impact on patients

Effectiveness

A large number of studies do not find any causal relationship between incentive payments and health outcomes (European Observatory in Health Systems (2014); Saint-Lary et al., (2012); Parkin E., 2018); Abellán et al., 2013; Fleetcroft et al ,2012; Bryan et al.,2001; Flodgren et al., 2011; Lavergne et al., 2016; Lavergne et al.,2018; Scott et al.,2011; Ryan et al., 2016; Gavagan, 2010; Houle et al.,2012; Fleetcroft, 2015; Fleetcroft, 2006; Cheryl et al.,2014; Guilliam, 2015; Wright, 2012). Other studies report outcome improvements in some process indicators (Patients Meeting Target, % Patients getting an specific treatment, etc..) (European Observatory in Health Systems,2014; Fleetcroft et al., 2012; Bryan et al., 2001, Guill P., 2015; Lavergne et al., 2016; Wright, 2012; Saint-Lary et al.,2012; Gavagan et al.,2010; Lehtovuori et al.,2015; Campbell et al., 2009; Johnson et al., 2018; Lin et al., 2016; Park et al., 2016; Guill, 2015) .There is evidence of the potential effectiveness of aligning payments with costs and quality in performance (Greene et al., 2015).

Equity

High scores in certain indicators may have been achieved by excluding patients from treatment or manipulating participant numbers, as P4P programs are often managed by medical professionals (Johnson et al.,2015; Guilliam et al.,2012; Scott et al., 2011). The presence, increase, or persistence of inequities appears to be independent of the health area (Chung et al., 2010; Hackett et al., 2014, Kontopanelis et al., 2016; Neil et al., 2015). However, healthcare regions with lower per capita income are more likely to exclude patients, thereby increasing inequities (Tim Doran et al., 2006; Guilliam et al., 2012; Kontopanelis et al., 2016). Additionally, P4P schemes that differentiate between various categories of patients and

providers may reduce salary disparities among professionals (Hackett et al.,2014). Patients suffering from conditions not incentivized under these schemes may receive suboptimal care42. Nevertheless, patients have not reported significant changes in perceived service quality since the introduction of the incentive systems (Shane et al., 2013; Shah et al., 2011; Lavergne et al., 2016).

Organization

Impact on organizational results

The organizational model can influence the outcomes of health services (Fleetcroft, 2015; Lehtovuori et al., 2015). The way healthcare providers manage resources and are compensated creates perceived incentives to improve patient care (Tim Doran et al., 2006; Shane et al., 2013; Steel et al., 2007). In this sense, incentives appear to have a positive effect when integrated into a broader quality improvement strategy, which also includes a feedback structure (Hearld et al., 2014; Saint-Lary et al., 2015; Greene J., 2013; Fernández Urrusuno et al., 2014); Campbell et al., 2010; LaaRocca et al., 2017; Johnson et al.,2015). Therefore, using mixed P4P strategies alongside other individual and group incentive models may be more effective than P4P schemes that rely on only one approach (Wright M.,2012; Lester et al.,2013); Johnson et al., 2018; Capizzi et al., 2011; Peckham et al., 2010; Scott et al.,2009; Johnson et al.,2018; Kolozsvari et al., 2014; Stewart et al., 2018; Marcotte et al., 2017; Greene et al.,2015; Neil et al.,2015). However, inefficiencies in payment systems can arise depending on the mechanism used, as simpler tasks are often favoured over complex ones, and there is a risk of duplication within the same process (Kecmanovic et al.,2015; Johnson et al.,2015).

Impact on satisfaction

The types of incentives pursued differ in the short and long term (Kirschner et al., 2012; Hannon et al.,2012; Mahmood, 2003; McDonald, 2009), and these may impact autonomy and professionalism(Scott et al., 2011; Kirschner et al.,2013; Alexander et al.,2013; Gemmill, 2008; Moreno et al.,2006; Gené-Badía et al.,2004; Ammt et al., 2017; Lester et al., 2013; Baek et al.,2013). Since the implementation of pay-for-performance, patients have not reported noticeable changes in clinical care (NHS quality Outcome Framework-Report of July 2018;

Olivier Saint Lary et al., 2015; Gené-Badía et al.,2004). It is suggested that the unintended consequences of P4P relate to the way these programs are designed and implemented (McDonald, 2009)

Unintended consequences

Unintended consequences vary and may depend on how incentives are implemented (McDonald et al., 2007; Kirschner et al., 2013; Bryan, 2001). At times, there is a disconnection between the indicators and the health outcomes they are meant to achieve (European Observatory in Health Systems, 2014). In other instances, statistical manipulation (gaming) can be observed, as well as an excessive focus on incentivized aspects (tunnel vision) (Gemmill, 2008). Additionally, the volume of prescription medications may also be impacted (McDonald et al.,2008); Gené-Badia et al.,2007), McDonald et al.,2008).The GPs are employing more staff and computerizing their clinical records so it can exist unintended consequences (Roland et al., 2006)

Withdrawal impact of indicators

Improvements in the quality of care slowed down after the first year of QOF (Wright, 2012); Johnson et al., 2018), and performance levels remained at the incentive level even after the withdrawal of incentives (Lin et al., 2016). The abolition of financial incentives was associated with reductions in recorded quality of care for most performance indicators (Daniel et al., 2023), Changes to pay-for-performance schemes should be carefully designed and implemented to monitor and address any declines in care quality (Saint-Lary et al., 2015); Kolozsvari et al., 2014); Houle et al., 2012); Green J., 2013); Capizzi, 2011), Friedberg et al.,2010), Lemak et al., 2013).

Cost-effectiveness

As Eijkenaar et al. (2013) assert: “Findings suggest that P4P can potentially be cost-effective, but the evidence is not convincing”). Therefore, there is no robust evidence supporting cost-effectiveness of P4P programmes at the moment. To improve health efficiently, the UK should redesign QOF or pursue alternative interventions according with (Pandya et al., 2018). To evaluate its cost-effectiveness, would require that new schemes be designed from the onset to

support their evaluation: control and treatment groups coupled with before and after data as said in Allen et al. (2014).

Discussion

The review presented in this article highlights both the strengths and weaknesses of the implementation and effects of P4P systems in the primary care sector. Let us first address the weaknesses and then provide some insights into how an incentive scheme can succeed, based on the strengths we have identified.

Firstly, there is no universally valid incentive scheme; the impact of P4P systems varies significantly depending on the country and the organizational environment in which they are implemented (European Observatory in Health Systems and Policies Series, 2014). Secondly, P4P systems may promote a minimum standard quality but only on processes in which monitoring indicators have been properly established. As Johnson et al. (2018) remark, both the size of incentives and the monitoring indicators should be carefully designed. Otherwise, there is a risk that no significant difference arises in relevant outcomes, such as mortality rate, between organizations incentivized through P4P systems and those that do not receive incentives. Thirdly, incentives might have counterproductive effects on practitioners' behaviour and the image that society has of doctors. In this way, professionals might be more prone to accomplish certain objectives because of the economic retribution they would receive. Moreover, they could abandon their mission within the system, practicing a sort of gaming, such as modifying the statistics to determine whether a goal has been reached for a specific incentivised indicator. Accordingly, health professionals could appear as "micro-bonuses' seekers", being perceived by patients as unmotivated or selfish.

On the contrary, if indicators are not chosen adequately, practitioners might not be involved with the targeted objectives and might perceive them as bureaucratic impositions, for they would be paying for processes that do not influence the results. Furthermore, there is a risk that once incentives are withdrawn, performance levels return to the original values (depending on the indicators or processes). Finally, it has also been found that economic incentives tend to be

demotivating in the long run if other organisational incentives do not accompany them. In fact, if incentives are established on the basis of the most efficient workers or centres' results, the rest of the staff might be encouraged to follow them rather than to focus on their own work quality.

In conclusion, we can assert that P4P systems entail some risks. It is important to highlight that there is currently limited evidence supporting improvements in coordination, cost-effectiveness, and continuity of care. Additionally, potential improvements are not guaranteed to be sustained over time, and patients with serious conditions may be negatively affected if doctors focus on generating statistics through achieving specific micro-results, potentially overlooking the overall health and well-being of patients.

However, the fact that P4P systems are exposed to some problems does not mean they should be discarded. The reviewed evidence offers valuable lessons that can help healthcare managers design effective incentive schemes. One important lesson is that economic incentives should not be the sole focus; they should be accompanied by other types of rewards. Economic incentives need to be combined with organizational incentives, promoting autonomy, work culture, and professional development. Additionally, the targets must be well aligned with clinical goals. As Bokhour et al. (2006) state: "If the quality targets are well aligned with professional clinical goals, then financial incentives may not be necessary. Performance monitoring and feedback, in the absence of financial rewards, may be enough to improve quality". To improve quality of care and reduce inequalities, both targets and initial values of the objective indicators should be appropriately set based on professionals' experience, allowing for sufficient room for improvement at their initial levels.

A second key is that indicators must be set both in absolute and relative terms, especially for processes and intermediate results, and adjusted for risk (i.e. patient complexity) when necessary. It is also required to modify the indicators periodically (for instance, annually), as well as avoiding introducing ones with contradictory objectives. The payment ought to be of an appropriate amount (without it being insignificant) and frequent over time.

Other two relevant lessons that we can cite are, on the one hand, that a close relationship among all the stakeholders (professionals, managers and patients) should be sought, providing them

with detailed and up-to-date information on the incentive programme, and on the other hand, that it is quite convenient to adapt the incentive scheme to the different generations that coexist within health organizations, since motivations are different depending on the professionals' age, gender and type of contract.

Lastly, our review points out that the success of any incentive scheme is determined by the organizational setting in which healthcare professionals carry out their work. Our study suggests that the most effective approach would be to establish a mixed model of incentives, incorporating both individual and collective elements, within an integrated management structure. This structure should involve health professionals in mixed management roles, where they act as managers, interact directly with suppliers, and participate in risk-sharing agreements. Such an organizational model would facilitate achieving the desired levels of autonomy for the staff and create an incentive scheme directly linked to their performance, fostering a sense of identification among the staff.

Obviously, our study is not without limitations. One of them arises from restricting our search primarily to studies published in journals. As a result, studies classified as part of the so-called 'grey literature' (with some exceptions) have largely been excluded. Nevertheless, we are confident that the most relevant studies have been included in our review. Another limitation stems from the fact that we only considered publications written in English or Spanish. Therefore, we cannot rule out the possibility that some important P4P experiences described in other languages may have been overlooked. Besides, our systematic search was conducted solely through MEDLINE, Scopus, and PubMed. While these are three of the most significant bibliographic databases, it is possible that they do not cover all relevant studies on the topic discussed in this article, and as such, some may not have been identified. Lastly, the systematic search ended in 2019, excluding therefore the pandemic years. This was done as an attempt to obtain evidence as much structural as possible. Although some studies considered particularly relevant published later have been also included in our review, we acknowledge that we may have overlooked some other important studies.

We believe that one of the main implications of our findings is that considering all the strengths and weaknesses of the reviewed P4P experiences could help to design and guide an optimal incentive scheme to enhance the performance of healthcare professionals working in the public primary care sector in Spain. As Gené-Badía (2021, p.2) states: 'We must take advantage of variable remuneration to encourage the development of new activities aimed at improving the quality of service from a primary care perspective.' We hope this paper has helped to identify at least one possible way to put Gené-Badía's words into action.

Conclusions

Throughout this thesis, three different challenges faced by the main stakeholders (managers, patients, practitioners) in the health system have been addressed. The first concerns how to improve the efficiency and equity with which available resources for reimbursing health technologies are allocated by decision-makers. The second challenge deals with valuing the utility of a 'process' attribute (i.e., attributes not strictly related to health but that also impact individual well-being) that has been scarcely studied: the information contained in medicine leaflets. The final challenge focuses on how to effectively incentivize the performance of health personnel working in Primary Care. Each of the chapters in this thesis is intended as a response to these challenges. Next, we briefly describe the methodology and main findings of each chapter, mention the unavoidable limitations, and discuss their implications for health policy, finally suggesting some directions for future research

First response: a framework for prioritizing health technologies

The first chapter of the thesis approaches the first of the mentioned challenges by developing a multiple criteria decision analysis (MCDA) framework to inform decisions on the incorporation of high-impact technologies in the Regional Health Service of the Region of Murcia (SMS).

The main conclusion of this study is the feasibility of developing a framework based on the MCDA methodology to guide purchasing decisions for new high-impact technologies in a Spanish regional health service, where no formal procedure with objective criteria currently exists for making such decisions. Moreover, this prioritization framework combines the

preferences of managers, healthcare professionals, and the general population, which we believe is one of the strengths of the study.

The methodology followed in our study to develop the intended framework encompassed two successive stages. In the first stage, key stakeholders within SMS, including clinical leaders and management personnel, participated in a focus group (n = 11) to discuss a list of proposed criteria provided by the research team. Fourteen criteria were selected from the latest version of the EVIDEM (Evidence and Value: Impact in Decision Making) framework (Evidem Collaboration, 2017). Additionally, criteria from the Belgian Health Care Knowledge Centre (KCE) framework (Polain et al., 2010) were integrated, with appropriate modifications when necessary. All criteria (sometimes divided into sub-criteria) were then grouped into five domains. This initial proposal was presented to the eleven members of the focus group responsible for selecting the final criteria. Domains, criteria, and sub-criteria were then voted on, followed by a debate and discussion of the results, which, if necessary, could lead to an extension or reduction of the domains. Ultimately, fifteen criteria emerged from this process, grouped into five dimensions (Need for intervention; Outcomes of the intervention; Knowledge about the intervention; Economic impact; Feasibility).

In the second phase, weights were assigned to the domains, criteria, and sub-criteria of the framework by two distinct samples: decision-makers and healthcare professionals from the SMS (n = 35), and a sample drawn from the general population of the Region of Murcia (n = 500). Participants in the first sample were surveyed online, whereas computer-assisted personal interviews (CAPI) were conducted in the homes of participants for the general population sample. To assign weights to the domains, criteria, and sub-criteria, each participant successively distributed 100 points among the domains, 100 among the criteria within each domain, and 100 among the sub-criteria within each criterion. Among the five domains, “Need for intervention” and “Outcomes of the intervention” were the most highly weighted by both samples. “Affected population,” “Disease severity,” and “Quality of the evidence” ranked at the top among the 15 criteria, a result similar to that found in other studies (Castro, Goetghebeur, and Moreno-Mattar, 2016; Iskrov and Stefanov, 2016; Mirelman et al., 2012). Although there is

a notable coincidence between the weights assigned by the general population and those assigned by decision-makers, some differences arise. Specifically, healthcare professionals place more importance on economic aspects, with the only significant discrepancy between the two samples observed in the domain "Economic impact." Additionally, responses from healthcare professionals were much more homogeneous than those from the general population. Given the different nature of preferences and the significant difference in sample sizes between the two surveys, we decided to use the average of the means obtained from the two samples for each item.

Notwithstanding, obtaining the intended framework does not mean that our study is free from limitations. One limitation stems from the method chosen (the 100-point allocation procedure) for weighting the criteria. Specifically, participants in both samples tended to distribute points roughly equally between criteria and sub-criteria, which may reflect the so-called equalizing bias (Rezaei, Arab, and Meheregan, 2022). Moreover, they also tended to use round numbers, as is common in this type of point allocation exercise (Honda, Kagawa, and Shirasuna, 2022). Despite these issues, the method we chose has the advantage of being simple and easy to understand. Another limitation of our study is the disparity in the sample sizes of both groups of respondents, which may help explain the differences in the degree of response dispersion. Without a doubt, it would have been desirable to have had more participation from SMS staff. Furthermore, we acknowledge as a potential limitation the omission of incorporating the general population's perspective in the initial phase of criterion identification, rather than only in the weighting phase. Lastly, no qualitative criteria were included in the framework, which could be seen as problematic if the score obtained with the multi-attribute function used to evaluate each high-impact technology becomes the only input considered for decision-making. For this reason, it is advisable that decision-makers do not rely on the resulting score automatically, without first following a deliberative process.

We believe that a direct outcome of implementing this framework would be increased transparency in prioritization decisions within the regional health system. Additionally, we think these decisions would become more deliberate and rational if decision-making bodies adopted

this framework. However, future research, after the framework has been in place for some time, could evaluate whether it has indeed been beneficial for decision-makers in the regional health service. Similarly, a new study could be conducted to incorporate patients' perspectives and examine whether they differ significantly from those of professionals, managers, and citizens.

Second response: an estimation of the monetary value of the information contained in medicine leaflets

The second study presented in this thesis applies the contingent valuation (CV) methodology to estimate the monetary value of the information contained in medicines leaflets. A direct antecedent to our work is that conducted by Dealy et al. (2021), who estimated the willingness to pay (WTP) for the information contained in traditional leaflets, without offering any quantitative information on efficacy and side effects. The novelty of our study is that it estimates the value of providing additional quantitative information on potential benefits and side effects of a hypothetical medicine, according to the best evidence available about risk communication.

The main conclusion of this study is that the access to information about the effectiveness and side effects of drugs holds inherent value for the population. The study indeed provides valuable insights into the WTP for quantitative information about drugs' benefits and side effects, displayed according to Yamagishi's (1997) recommendations for effective risk communication.

A CV survey was designed to estimate the value of the additional information contained in a complementary brochure to the usual patient information leaflet (PIL) of a hypothetical medicine. This medicine was an anticoagulant indicated for preventing cardiovascular diseases. A sample of 217 adults selected to represent the age and gender distribution within the general population were randomly assigned to two distinct groups (n1 = 110; n2 = 107). The questionnaires administered to the subjects in each of the groups were identical, except for the section inquiring about their WTP.

Prior to asking for WTP, two different, though complementary, pieces of information on the anticoagulant medicine and its consequences were presented to the participants. The first piece of information was a "traditional" leaflet, and the latter one a brochure providing additional

quantitative information on the effectiveness and side effects of the medicine. As noted above, the design of this complementary brochure responds to best practices on risk communication reported in the literature, particularly Yamagishi's (1997) recommendations.

Once both the PIL and the complementary brochure were shown to the respondents, they were asked to state their maximum WTP for the medicine in different ways depending on the group to which they belonged. In Group 1, participants first stated their WTP for the medicine whose package only included the traditional leaflet, and then stated their WTP for also including the brochure containing detailed information on the anticoagulant's benefits and harms within the package. Thus, they valued the medicine and the complementary information provided by the brochure separately. In contrast, in Group 2, participants stated their maximum WTP for the medicine package as a whole, including both the PIL and the complementary brochure. Afterwards, respondents belonging to Group 2 were asked to indicate the proportion (as a percentage) of the total WTP they had previously stated that was attributed exclusively to the additional information provided in the brochure.

The main finding of our study is that the WTP for the additional information included in a brochure, complementary to a traditional PIL, ranged from 60 cents to 1 euro per month. Interestingly, our upper value closely aligns with the average WTP ($ 1.37) for standardized informational leaflets reported by Dealy et al. (2021) for the United States. Our results also clearly indicate that the elicitation format used to estimate the WTP for the brochure is not irrelevant. In this sense, the mean WTP for the brochure in Group 2 exceeded nearly 60% of the mean WTP elicited in Group 1, even though around one third of participants in Group 2 assigned a zero WTP value to the brochure.

Some possible explanations for the elicitation effects recorded are provided in chapter 2. One of them is imprecision in preferences (MacCrimmon and Smith, 1986). As WTP for the brochure in Group 2 was inferred from the percentage rated onto a 0-100 visual analogue scale (VAS), if preferences are imprecise then many respondents in Group 2 would be not sure about the exact percentage to attribute to the brochure, easily giving rise to values higher than those stated in Group 1 (elicited by means of a payment card followed by an open-ended question). Another

reason for the sharp asymmetry between the WTP for the brochure of the two groups could be attributed to respondents' attention (Bordalo et al., 2013, 2016). In this way, whereas in questionnaire 1, once respondents have stated their WTP for the medicine and before they were asked to state their WTP for the information, they were newly informed of the utility of the brochure to be more awareness of the benefits and risks of the medicine intake, in questionnaire 2, however, immediately after of stating the WTP for the whole medicine package, and without further elaboration, respondents were inquired about the percentage they would place on the brochure. In this regard, probably respondents' attention was directed to the medicine and not to the brochure, causing many of them underweighted the value of the additional information providing a zero WTP value.

Our study has several limitations that warrant consideration. Firstly, while we believe that the sample size used was sufficient for the specific analysis conducted, we acknowledge that in order to generalize our results to the broader adult Spanish population, a larger and more diverse sample should have been employed. This is an issue that should be addressed in future analyses. Similarly, our study does not provide insights on patients' willingness to pay (WTP) for information in real clinical settings. This is another aspect that would be worth exploring in future research. Lastly, a further limitation of the study is the asymmetry in how WTP for additional information was assessed between the two respondent groups. Therefore, it would be valuable to investigate the extent the findings reported here might change if participants in Group 2 were informed about the benefits and risks of the medication before being asked to attribute a percentage of their WTP for the medicine to the brochure.

The findings presented in this study suggest the need to revise the manner in which information about the effectiveness and side effects of medicines is typically conveyed. Our research shows that providing quantitative data on benefits and risks in medicine leaflets holds significant value for people. Such data cannot be displayed in just any way; rather, they should follow the best available evidence on effective risk communication.

Third response: pay-for-performance in Primary Care

In the third chapter of this thesis some key features that an incentive scheme should have to effectively improve primary care professionals' performance in the Spanish National Health Service are shown. Such features are identified as a result of a systematic review of articles and reports published between 1999 and 2019 (supplemented with a few later publications regarded particularly relevant).

A remarkable conclusion of this study is that the evidence showing that pay-for-performance (P4P) systems entail some risks does not mean they must be discarded. As our review suggests, there are several critical properties that, if verified, could enhance the effectiveness of the incentive system. Our analysis suggests that the most effective approach to ensure the success of a P4P model would be to establish a mixed model of incentives, incorporating both individual and collective elements, within an integrated management structure.

A systematic search of published studies analyzing the influence of P4P schemes on primary care outcomes was conducted using three bibliographic databases (MEDLINE, Scopus, and PubMed) from 1999 to 2019. The pandemic years were set aside because throughout, incentive systems suffered many changes in order to ease the concentration of resources on responding to the pandemic challenge. Nevertheless, some subsequent studies, considered particularly relevant, were also included in our review.

The selection of articles for this review adhered to the PRISMA 2020 reporting guidelines for systematic reviews and meta-analyses. A total of 108 articles (102 from the systematic review and 6 from other methods) were chosen from the 773 initially identified. Next, the selected studies are classified according to five criteria: impact on patients, organization, unintended consequences, withdrawal of indicators, and cost-effectiveness.

The evidence collected through the five precedent criteria shows lights and shadows. Starting with the latter, our study demonstrates that there is currently limited evidence supporting improvements in coordination, cost-effectiveness, and continuity of care. Additionally, potential improvements are not guaranteed to be sustained over time, and patients with serious conditions

may be negatively affected if doctors focus on generating statistics through achieving specific micro-results, potentially overlooking the overall health and well-being of patients.

However, our study also finds five features that can help an incentive scheme succeed. Firstly, economic incentives need to be combined with organizational incentives, promoting autonomy, work culture, and professional development. Secondly, performance indicators must be set both in absolute and relative terms, especially for processes and intermediate results, and adjusted for risk (i.e. patient complexity) when necessary. In addition, a close relationship among all the stakeholders (professionals, managers and patients) should be sought, providing them with detailed and up-to-date information on the incentive scheme. Fourth, it is quite convenient to adapt the incentive scheme to the different generations that coexist within health organizations. Lastly, it seems that that the success of any incentive scheme is determined by the organizational setting in which healthcare professionals carry out their work.

Obviously, our study is not without limitations. One of them arises from restricting our search primarily to studies published in journals. As a result, studies classified as part of the so-called 'grey literature' (with some exceptions) have largely been excluded. Another limitation stems from the fact that we only considered publications written in English or Spanish. Besides, our systematic search was conducted solely through MEDLINE, Scopus, and PubMed. Lastly, the systematic search ended in 2019, excluding therefore the pandemic years. In summary, although we are convinced that the inclusion/exclusion criteria used are reasonable, we cannot be completely sure that we have not overlooked any valuable studies.

To us the main implication of our findings is that the adherence to the strengths (and the avoidance of the weaknesses) of the reviewed P4P experiences could help to design and guide an optimal incentive scheme to enhance the performance of healthcare professionals working in the public primary care sector in Spain.

A final corollary

There is a clear throughline that underlies this thesis: the aim of improving the public health system at several levels. Several proposals are provided in an attempt to fulfill that goal. These proposals are driven by the application of different methodologies. In this way, Multi-Criteria Decision Analysis (MCDA) is presented as a tool that can help decision-makers to set priorities in the incorporation of new health technologies. As demonstrated, it is feasible to endow a regional health system with a MCDA framework that can be used to inform decisions on the funding of high-impact technologies. Similarly, Contingent Valuation is used in this thesis as a methodology able to estimate the value that citizens give to a more detailed information on benefits and risks of medicines. The fact that the population is willing to pay an additional amount of money for the inclusion of this information opens a line of research that could lead to the transformation of patient information leaflets as we know them today. Finally, a systematic review of pay-for-performance (P4P) experiences endorsed by the PRISMA reporting guidelines is conducted. Its results depict the weaknesses and strengths of P4P models whose insights can contribute to designing and implementing better incentives in the healthcare system.

The author of this thesis is convinced that reality is complex and that the challenges faced in the healthcare sector cannot be addressed with quick fixes, intuition, or by relying solely on experience. Solid analytical procedures grounded in scientific evidence must be put into practice. I hope this thesis serves as a good example of that.

Appendix 5

Figure A 9. Conceptual Framework

This conceptual framework is a visual representation that helps to illustrate the expected relationship between a change into the parameters of healthcare economic incentives, through a pay for performance system, and improvements in the quality of healthcare and its potential results.

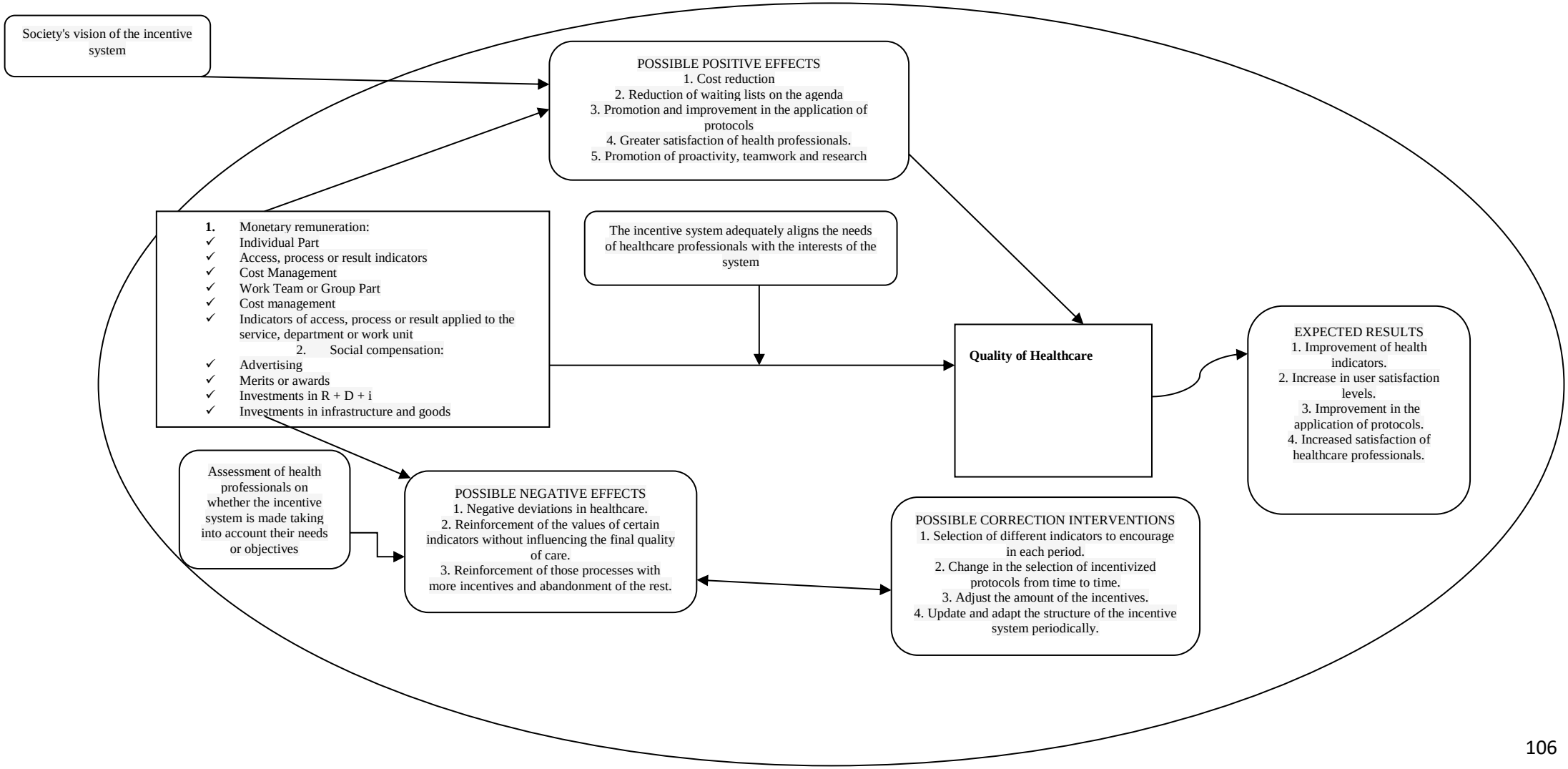


Figure A 10.PICO Analysis (Population, Intervention, Comparison, Outcome)

- What characterises the population of this study?

The main population are the healthcare professionals

- What is the condition or subject you are interested in?

How a pay for performance incentive scheme may influence the way these professionals act and if it is possible to affect the quality of healthcare.

1. INTERVENTION OR EXPOSURE

- What do you want to do with this population?

To implement a P4P scheme and analyse whether healthcare results as well as patients and professionals satisfaction increase with it.

2. COMPARISON

- What is the alternative to the intervention?

A mix of different kind of incentives (social and economics), no incentive scheme.

3. OUTCOME

- What are the relevant outcomes?
- Reduction of costs, Improvement of the healthcare results, Reduction of waiting lists, Improvement of Teamwork and satisfaction of professionals.

QUESTION: *"Is there an incentive scheme based on pay for performance with satisfactory outcomes in healthcare management and positive overall results?"*

5.REFERENCES

- Abellán Perpiñán, Jose Maria; Martinez Pérez, Jorge Eduardo;Sánchez Martínez, Fernando Ignacio;Serrano Martínez, Lorenzo;Méndez Martínez, Ildefonso (2013).The public health system in Spain and its autonomous communities. Sustainability and reforms. BBVA Foundation. ISBN 978-84-92937-45-5
- Alexander J.A., Maeng D.,CasalinoL.P. (2013). Use of care management practices in small and medium sized physician groups: Do public reporting o physicians quality and financial incentives matter? *HealthServicesResearch*, *48*((2 PART 1)), 376–97. DOI: https://doi.org/10.1111/j.1475-6773.2012.01454.x
- Allen T., Whittaker W., M. T. (2014). Impacts of pay for performance on the quality of primary care. *Risk Management and Healthcare Policy*, *7*, 113–120. DOI: https://dx.doi.org/10.2147%2FRMHP.S46423
- Allen T., Whittaker W., Sutton M., K. E. (2018). Influence of financial and reputational incentives on primary care performance: A longitudinal study. *British Journal of General Practice*, *68*(677), e811–18. DOI: https://doi.org/10.3399/bjgp18X699797
- Ammt M., F. G. (2017). The influence of welfare systems on pay-for-performance programs for general practitioners: A critical review. *Social Science& Medicine*, *178*, 157–66.
- Ankur Pandya, Tim Doran, Jinyi Zhu, et al.Modelling the cost-effectiveness of pay-for performance in primary care in the UK (2018) .BMC Medicine.2018;16:135. DOI:https://doi.org/10.1186/s12916-018-1126-3
- Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review*, 53(5), 941-973.
- Ash A.S., E. R. P. (2012). Risk-adjusted payment and performance assessment for primary care. *Medical Care*, *50*(8), 643–53. DOI: https://dx.doi.org/10.1097%2FMLR.0b013e3182549c74

- Ashworth, Mark & Lea, Robert & Gray, Heather & Gravelle, Hugh & Majeed, Azeem. (2003).The development of prescribing incentive schemes in primary care: A longitudinal survey. *The* British *journal of general practice: the journal of the Royal College of General Practitioners. 53. 468-70.*
- Australian National Audit Office(ANAO).Report of February(2007).Introduction; [consulted on 14/08/2019].Available in:https://www.anao.gov.au/work/performance-audit/application-outcomes-and-outputs-framework
- Baek J.D., Xirasagar S.,Stoskopf C.H. SRL (2013). Physician-targeted financial incentives and primary care physicians self-reported ability to provide high-quality primary care. J Prim Care Community Health. 2013;4(3):182–8.
- Baltagi BH, Lagravinese R, Moscone F, Tosetti E. Health care expenditure and income: a global perspective. *Health Econ*. 2017;26:863-74.
- Baltussen, R., & Niessen, L. (2006). Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation, 4*(1), 14.
- Baltussen, R., Marsh, K., Thokala, P., Diaby, V., Castro, H., Cleemput, I., Garau, M., Iskrov, G., Olyaeemanesh, A., Mirelman, A., Mobinizadeh, M., Morton, A., Tringali, M., van Til, J., Valentim, J., Wagner, M., Youngkong, S., Zah, V., Toll, A., Jansen, M., ... Broekhuizen, H. (2019). Multicriteria Decision Analysis to Support Health Technology Assessment Agencies: Benefits, Limitations, and the Way Forward. *Value in Health, 22*(11), 1283–1288.
- Basu, A., & Meltzer, D. (2005). Value of Information on Preference Heterogeneity and Individualized Care. *Medical Decision Making, 25*(6), 614-627.
- Bateman, I. Cameron, M. P., Tsoumas, A. (2006), "Investigating the characteristics of stated preferences for reducing the impacts of air pollution: a contingent valuation experiment". Working Papers in Economics 06/08, University of Waikato, New Zealand.
- Bateman, I., Day, B., Loomes, G. and Sugden, R. (2007), "Can ranking techniques elicit robust values?", *Journal of Risk and Uncertainty*, 34: 49–66.

- Bateman, I.J. et al. (2002), *Economic Valuation with Stated Preference Techniques: A Manual*, Cheltenham, United Kingdom: Edward Elgar.
- Bateman, I.J., Munro, A., Rhodes, B., Starmer, C., Sugden, R. (1997), "Does part-whole bias exist? An experimental investigation", *Economic Journal*, 107: 322–332.
- Belton, V., & Stewart, T. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer Science & Business Media.
- Benavent J., Juan C.,Clos J.et al(2009). Using pay for performance to introduce changes in primary healthcare centres in Spain: First year results. Qual Prim Care. 2009; 17(2):123–31.
- Bishop, R. C., Champ, P.A. and Mullarkey, D. J. (1995), "Contingent Valuation" en D. W. Bromley (ed.), *Handbook of Environmental Economics*, Cambridge, MA: Blackwell, 629–654.
- BLACK, G. (1982). Inequalities in Health: Black Report, serie: Pelican, Harmondsworth; New York, Penguin Books.
- Bokhour B.G.,Burgess Jr. J.F.,Hook J.M. et al(2006). Incentive implementation in physical practices: A qualitative study of practice executive perspectives on pay for performance. Med Care Res Rev. 2006;63(1):73S-95S.
- Bonfrer I, Figueroa J F, Zheng J, Orav E J, Jha A K. Impact of Financial Incentives on Early and Late Adopters among US Hospitals: observational study BMJ 2018; 360 :j5622 doi:10.1136/bmj.j5622.
- Bordalo, P., Gennaioli, N., Shleifer, A. (2013), "Salience and Consumer Choice", *Journal of Political Economy*, 121: 803–843
- Bordalo, P., Gennaioli, N., Shleifer, A. (2016), "Competition for Attention", *The Review of Economic Studies*, 83: 481-513.
- Briggs, A. (1961). The Welfare State in Historical Perspective. *European Journal of Sociology, 2*(2), 221-258.
- Briggs, A., Sculpher, M., & Claxton, K. (2006). *Decision Modelling for Health Economic Evaluation*. Oxford University Press.

- Broekhuuizen H, Groothuis-Oudshoom CGM, van Til JA, Hummel JM, Ijzerman MJ. A review and classification of approaches for dealing with uncertainty in Multi-Criteria Decision Analysis for healthcare decisions. *Pharmacoeconomics*. 2015;33:44555.
- Bryan S. Armor (2001).The effect of explicit financial incentives on physician behaviour. Arch Intern Med.2001; Vol 161, p.1261-66.
- Butler, D. J., Loomes, G. C. (2007), "Imprecision as an account of the preference reversal phenomenon", *American Economic Review*, 97(1): 277–297.
- Campbell S.,Reeves D.,Kontopantelis E.,Middleton E., Sibbald B.,Roland M.(2009). Quality of primary care in England with the introduction of pay for performance. N Engl J Med. 2009;357(2):181-9. DOI: 10.1056/NEJMsr065990
- Campbell S.M., Scott A., Parker R.M. , Nacarella L.,S. Furler J.,Young D., M. Shivey P.,(2010). Implementing pay for performance in Australian primary care: Lessons from the United Kingdom and the United States. *Medical Journal of Australia*, *193*(7), 408–11. https://doi.org/10.5694/j.1326-5377.2010.tb03971.x
- Campbell, S. M., Reeves, D., Kontopantelis, E., Sibbald, B., & Roland, M. (2009). Effects of pay for performance on the quality of primary care in England. *The New England Journal of Medicine*, *361*(4), 368–378.
- Campillo-Artero C, Puig-Junoy J, Culyer A. Does MCDA trump CEA? *Appl Health Econ Health Policy*. 2018;16(2):147-51.
- Campolina AG, Estevez-Diz MDP, Abe JM, de Soárez PC. Multiple Criteria Decision Analysis (MCDA) for evaluating cancer treatments in hospital-based health technology assessment: The Paraconsistent Value Framework. *PLoS One*. 2022;17:e0268584.
- Cantarero Prieto, D. (2008). Desigualdades territoriales y equidad en el sistema sanitario español. *Cuadernos económicos de ICE*, *75*, 141-166.
- Capizzi, Silvio, Specchia M.L.,Veneziano M.A., Ricciardi W. (2011). Pay for performance in primary care. *Igiene e Sanita Pubblica Sanita Publ*, *67*(5), 531–40.

- Caro A, Valcárcel MC, Olry A. Valor de la hemodiálisis concertada y la hospitalaria mediante un análisis de decisión multicriterio. *Nefrología*. 2022 Sep 17; 10.1016/j.nefro.2022.08.004.
- Carson, R. T., Mitchell, R. C. (1995), "Sequencing and nesting in contingent valuation surveys", *Journal of Environmental Economics and Management*, 28: 155–173.
- Carson, R.T., (1997). "Contingent valuation surveys and tests of insensitivity to scope", en R.J. Kopp, W. Pommerhene, and N. Schwartz (eds), *Determining the Value of Non-marketed Goods*. Boston: Kluwer Academic Publishers, 127–163.
- Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N. and Spencer, A. (1999), "On the contingent valuation of safety and the safety of contingent valuation: Part 2-The CV / SG "Chained" approach", *Journal of Risk and Uncertainty*, 17(3): 187-213.
- Castro HE, Goetghebeur M, Moreno-Mattar O. Testing Multi-Criteria Decision Analysis for More Transparent Resource-Allocation Decision Making in Colombia. *Int J Technol Assess Health Care*. 2016; 32:307-14.
- Chaix-Couturier C., Durand-ZaleskiI.,Jolly D (2000). Effects of financial incentives on medical practice: Results from a systematic review of the literature and methodological issues. Int J Qual Heal Care. 2000;12(2):133–42.
- Champ, P.A., Bishop, R.C. (2006), "Is Willingness to Pay for a Public Good Sensitive to the Elicitation Format?", *Land Economics* 82(2): 162–173.
- Cheryl Cashin, Y-Ling Chi, Peter Smith MB and ST(2014). Paying for performance in health care. Implications for health system performance and accountability. Cheryl Cashin, Y-Ling Chi, Peter Smith MB and ST, editor. Open University Press European Observatory on Health Systems and Policies Series; 2014. 312 p.
- Chung S., Palaniappan L.P., Trujillo L.M. et al (2010). Effect of physician-specific pay-for-performance incentives in a large group. *The American Journal of Managed Care*, *16*(2), e35-42. https://dx.doi.org/10.1111%2Fj.1475-6773.2009.01072.x

- Clark, J., Friesen, L. (2008), "The causes of order effects in contingent valuation surveys: An experimental investigation", *Journal of Environmental Economics and Management*, 56: 195-206.
- Claxton K. Three questions to ask when examining MCDA. Value outcomes spotlight 2015;January/February:18-20.
- Cleemput I, Devriese S, Kohn L, Devos C, van Til J, Groothuis-Oudshoorn K, Van der Voorde C., Vandekerckhove P. *Incorporating societal preferences in reimbursement decisions - relative importance of decision criteria according to Belgian citizens*. KCE reports. Brussels: Belgian Health Care Knowledge Centre (KCE); 2014.
- Cohen G.R.,Erb N., L. C. H. (2012). Physician practice responses to financial incentive programs: Exploring the concept of implementation mechanisms. *Advances in Health Care Management*, *13* (article number 17058385), 29–58.
- Cokely E. T., Ghazal S., García-Retamero R. (2014), "Measuring numeracy" en B. L. Anderson, J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health*. Cambridge, UK: Cambridge University Press.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., Garcia-Retamero, R. (2012), "Measuring risk literacy: The Berlin numeracy test", *Judgment and Decision Making*, 7(1): 25–47.
- Comín, F. (1996). Las formas históricas del Estado de Bienestar: el caso español. En *Dilemas del Estado de Bienestar*. Fundación Argentaria.
- Connelly L.B., B. J. R. G. (2012). Insurance rebates, incentives and primary care in Australia. *Geneva Papers on Risk and Insurance: Issues and Practice*, *37*(4), 745–762
- Cookson, R., Mirelman, A. J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O. F., Verguet, S., & Culyer, A. J. (2017). Using Cost-Effectiveness Analysis to Address Health Equity Concerns. *Value in Health*, *20*(2), 206–212.
- Morales D R, Minchin M, Kontopantelis E, Roland M, Sutton M, Guthrie B et al. Estimated impact from the withdrawal of primary care financial incentives on

selected indicators of quality of care in Scotland: controlled interrupted time series analysis BMJ 2023; 380 :e072098 doi:10.1136/bmj-2022-072098

- Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios CM, et al. Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. *Int J Technol Assess Health Care*. 2011;27:369-75.
- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big Data in Healthcare: Management, Analysis and Future Prospects. *Journal of Big Data, 6*(54), 1-25.
- De Montis A, De Toro P, Droste-Franke B, Omann I, Sagl S. Assessing the quality of different MCDA methods. In: Getzner M, Spash C, Stagl S, editors. Alternatives for environmental valuation. London, UK: Routledge; 2004.
- Dealy, B. C., Horn, B.P, Bohara, A.K., Berrens, R.P. and Bryan, A.D. (2017). "The Impact of Behavioural Risk-reduction Interventions on Willingness to Pay to Avoid Sexually Transmitted Infections: A Stated Preference Study of Justice-involved Youth", *Applied Economics*, 49(56): 5673–5685.
- Dealy, B.D., Kearsley, A., Wolff, C., Botkins, E., Lew, N. and Nardinelli, C. (2021), "Willingness to pay to standardize patient medication information", *Applied Economics*, 53(9): 1112-1126.
- Desvousges, W., Mathews, K., Train, K. (2016), "Reply to On the adequacy of scope test results: comments on Desvousges, Mathews, and Train", *Ecological Economics*, 130: 361–362.
- Devlin, N. & Sussex, J. (2011). *Incorporating multiple criteria in HTA. Methods and processes*. London: Office of Health Economics.
- Diamond, P.A., Hausman, J.A., Leonard, G.K. and Denning, M.A. (1993), "Does contingent valuation measure preferences? Some experimental evidence", en J.A. Hausman (ed.), *Contingent valuation: A critical assessment*. Amsterdam: North Holland Press.

- Dolan JG. Medical decision making using the analytic hierarchy process: choice of initial antimicrobial therapy for acute pyelonephritis. *Med Decis Making*. 1989;9:51–6.
- Dolan JG. Shared decision-making–transferring research into practice: the analytic hierarchy process (AHP). *Patient Educ Couns*. 2008;73:418-25.
- Donaldson, C., & Shackley, P. (1997). Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social Science & Medicine (1982), 44*(5), 699–707.
- Doran T, Fullwood C, Kontopantelis E, Reeves D (2008). Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. Lancet. 2008 Aug 30;372(9640):728-36. DOI: 10.1016/S0140-6736(08)61123-X. Epub 2008 Aug 11. PMID: 18701159.
- Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U., & Roland, M. (2006). Pay-for-performance programs in family practices in the United Kingdom. *The New England Journal of Medicine, 355*(4), 375–384.
- Dormont B, Grignon M, Huber H. Health Expenditure Growth: reassessing the threat of ageing. *Health Econ*. 2006;15:947-63.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the Economic Evaluation of Health Care Programmes* (4th ed.). Oxford University Press.
- Drummond, M. F., Stoddart, G. L., & Torrance, G. W. (1987). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.
- Dubourg, W.R., Jones-Lee, M.W., Loomes, G. (1994), "Imprecise Preferences and the WTP-WTA Disparity.", *Journal of Risk and Uncertainty* 9: 115–133.
- Dubourg, W.R., Jones-Lee, M.W., Loomes, G. (1997), "Imprecise preferences and survey design in contingent valuation", *Economica* , 64: 681-702.

- Educainee (2013), "Programa Internacional para la Evaluación de las Competencias de la población adulta (PIAAC) I", Boletín de Educación INEE Nº 15, Ministerio de Educación, Cultura y Deporte, 8 Octubre 2013.
- Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Polity Press.
- European Observatory in Health Systems and Policies Series. (2014). *Paying for Performance in Health Care: Implications for Health System Performance and Accountability*. OECD Publishing.
- EVIDEM Collaboration. Concept & definitions (v4.0). 10th Edition. 2017.
- Fanshel, S., Bush., J. W. (1970). A health status index and its application to health services outcomes. *Op. Res.* 18:1021.
- Fernández Urrusuno R.,Pérez Pérez P., Montero Balosa M.C, C. Márquez Calzada & B. Pascual de la Pisa (2014). Compliance with quality prescribing indicators linked to financial incentives: What about not incentivized indicators?: An observational study. *European Journal of Clinical Pharmacology*, *70*(3), 303–11. https://doi.org/10.1007/s00228-013-1610-9
- Fichera E. PM (2017). Pay for performance and contractual choice: the case of general practitioners in England. Health Econ Rev. 2017;7(1):Article Number 6.
- Fischhoff, B. (1995). Risk perception and communication unplugged: Twenty years of process. *Risk Analysis*, *15*(2), 137-145.
- Fleetcroft R (2015). The mortality rate in England between 2007 and 2012 was not associated with the quality of primary care in an established pay for performance programme. BMJ Evidence-BasedMed [Internet]. 2015 [consulted on 21/12/2018];20(149). http://dx.doi.org/10.1136/ebmed-2015-110205.Disponible en: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.953.8752&rep=rep1&type=pdf
- Fleetcroft R. (2006). Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Services Research and Policy*, *11*(1), 27–31. https://doi.org/https://doi.org/10.1258%2F135581906775094316

- Fleetcroft R., Parekh-Bhurke S., Howe A., Cookson R., Swift Louise, Steel N. (2010). The Uk pay for performance programme in primary care: Estimation of population of mortality reduction. Br J Gen Pract. 2010; 60(578):e345–52. DOI: https://doi.org/10.3399/bjgp10x515359
- Fleetcroft, R, Steel, N, Cookson, R, Walker S., Howe A. (2012). Incentive payments are not related to expected health gain in the pay for performance scheme for UK primary care: cross sectional analysis.BMC Health Services Research 2012; 12: 94. Available in: https://bmchealthservres.biomedcentral.com/track/pdf/10.1186/1472-6963-12-94
- Frank Eijkenaar, Martin Emmert, Manfred Scheppach, Oliver Schöffski (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. Health Policy, Volume 110, Issues 2–3, 2013, Pages 115-130, ISSN 0168-8510, DOI: https://doi.org/10.1016/j.healthpol.2013.01.008(https://www.sciencedirect.com/science/article/pii/S0168851013000183)
- Friedberg M.W., Safran D.G., Coltin K. et al (2010). Paying for performance in primary care: Potential impact on practices and disparities. *Health Affairs*, *29*(5), 926–932. DOI: https://doi.org/10.1377/hlthaff.2009.0985
- García-Altes, A., Rodriguez, M., Borrell, C., Pérez, G. (2008). Desigualdades en salud y en la utilización y el desempeño de los servicios sanitarios en las Comunidades Autónomas. *Cuadernos económicos de ICE* 75,115.128.
- Garcia-Retamero, R. and Galesic, M. (2011), "Using plausible group sizes to communicate information about medical risks", *Patient education and counseling*, 84(2): 245-250.
- Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., Traczyk, J. (2019), "Numeracy and Risk Literacy: What Have We Learned so Far?", The Spanish Journal of Psychology, 22, e10: 1–11.
- Gavagan T.F.,DuH.,Saber B.G. et al(2010). Effect of financial incentives on improvement in medical quality indicators for primary care. J Am Board Fam Med. 2010;23(5):622–31.

- Gemmell I, Campbell S, Hann M, Sibbald B. (2009). Assessing workload in general practice in England before and after the introduction of the pay-for-performance contract. J Adv Nurs. 2009 Mar; 65(3):509-15. doi: 10.1111/j.1365-2648.2008.04902.x. PMID: 19222648.
- Gemmill M (2008). Pay-for-Performance in the US: What lessons for Europe? *Eurohealth*, 13(4): 21-23.
- Gené-Badia J. (2021). Let's take advantage of the COVID-19 pandemic to overcome inefficient variable remuneration linked to performance. Aten Primaria. 2021; 53(2). DOI: DOI: https://doi.org/10.1016/j.aprim.2020.12.002
- Gené-Badía J, G. de P. P. (2004). Variable payment linked to quality of care. *Atención Primaria*, *34*(4), 198–201. https://doi.org/10.1016/s0212-6567(04)78908-8
- Gené-Badia J.,Escaramis-Babiano G.,Sans-Corrales M, Sampietro Colom L.,Aguado-Menguy F.,Cabezas-Peña C., Gallo de Puelles P. (2007). Impact of economic incentives on quality of professional life and on end user satisfaction in primary care. Health Policy (New York). 2007;80(1):2–10.
- Gerd Flodgren, Martin P Eccles, Sasha Sepperd, Anthony Scott, Elena Parmelli, Fiona R. Beyer, (2011). An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes.Cochrane DatabaseSystRev.2011 ; (7): CD009255. doi:10.1002/14651858.CD009255
- Gérvas, J., Pérez Fernández, M. Pago por desempeño («pay for performance» [P4P]). AMF.2008; 4(6): 344-52
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53-96.
- Gilabert-Perramon A, Torrent-Farnell J, Catalan A, Prat A, Fontanet M, Puig-Peiró R,Merino Montero S., Khoury Hanane, Mireille M. Goetghebeur, Xavier Badia al. Drug evaluation and decision making in Catalonia: development and validation of a

methodological framework based on multi-criteria decision analysis (MCDA) for orphan drugs. *Int J Technol Assess Health Care*. 2017;33:111-20.

- Gilliam S.J., Siriwardena A.N, Steel N. (2012). Pay for performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. *AnnalsFamilyMed, 10*(5), 461–68. https://doi.org/10.1370/afm.1377
- Glaize A, Duenas A, Di Martinelly C, Fagnot I. Healthcare decision-making applications using multicriteria decision analysis. A scoping review. *J Multi-Crit Decis Anal*. 2019;26:62-83.
- Goetghebeur M, Wagner M, Khoury H, Rindress D, Grégoire JP, Deal Ch. Combining multi-criteria decisión analysis, ethics and health technology assessment: applying the EVIDEM framework decision making framework to growth hormone for Turner syndrome patients. *Cost Effec Resourc Alloc*, 2010;8:4.
- Goetghebeur MM, Wagner M, Khoury H, Levitt R. J., Erikson Lonny J., Rindress D. (2008). Evidence and value: impact on Decision Making–the EVIDEM framework and potential applications. *BMC Health Serv Res*. 2008;8:270. doi: 10.1186/1472-6963-8-270
- Goetghebeur, M. M., Wagner, M., Khoury, H., Levitt, R. J., Erickson, L. J., & Rindress, D. (2012). Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Medical Decision Making, 32*(2), 376–388.
- Grant, S. and Huby, G. and Watkins, F. and Checkland, K. and McDonald, R. and Davies, H. and Guthrie, B. (2009).The impact of pay for-performance on professional boundaries in UK general practice: an ethnographic study. Sociology of Health and Illness, 31 (2). pp. 229-245. ISSN 0141-9889
- Greene J.,Kurtzman E.T.,Hibbard J.H., O. V. (2015). Working under a clinic-level quality incentive: primary care clinician's perceptions. *Annals of Family Medicine, 13*(3), 235–41. DOI: https://doi.org/https://dx.doi.org/10.1370%2Fafm.1779

- Greene, J. (2013). An examination of pay-for-performance in general practice in Australia. *Health Service Research*, *48*(4), 1415–32. DOI: https://doi.org/10.1111/1475-6773.12033
- Greve, B. (2021). *Handbook on Austerity, Populism and the Welfare State*. Edward Elgar Publishing.
- Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, *80*(2), 223-255.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393-408.
- Guill, P. (2015). Pay-for-performance and primary care physicians: lessons from the UK Quality and Outcomes Framework for local incentive schemes. *Journal of the Royal Society of Medicine*, *108*(3), 80–2. https://doi.org/10.1177/0141076815576701
- Guillén, A. M., & León, M. (Eds.). (2011). *The Spanish Welfare State in European Context*. Ashgate.
- Guilliam S. (2015). Financial incentive schemes in primary care. *Journal of Healthcare Leadership*, *7*, 75–80. https://doi.org/https://doi.org/10.2147/JHL.S64365
- Günther OH, Kürstein B, Riedel-Heller SG, König HH (2010). The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. *Health Serv Res. 2010 Feb;45(1):212-29*. doi: 10.1111/j.1475-6773.2009.01045.x. Epub 2009 Sep 24. PMID: 19780851; PMCID: PMC2813445.
- Haab, T.C., Interis, M.G., Petrolia, D.R. and Whitehead, J.C. (2013), ·From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation", *Applied Economic Perspectives and Policy*, 35(4): 593–612.
- Hackett J.,Glidewell L.,West, R., Carder P.,Doran T., Foy R. (2014). "Just another incentive scheme" a qualitative interview study of a local pay for performance scheme for primary care. *BMC Family Practice*, *15*(168). https://doi.org/10.1186/s12875-014-0168-7

- Hamrosi, K., Dickinson, R., Knapp, P., Raynor, D., Krass, I., Sowter, J. and Aslani, R. (2013), "It's for your benefit: exploring patients' opinions about the inclusion of textual and numerical benefit information in medicine leaflets", *International Journal of Pharmacy Practice*, 21: 216-225.
- Hannon K, L., Lester H.E. CSM (2012). Patient's views of pay for performance in primary care: A qualitative study. Br J Gen Pract. 2012;62(598):e322–8.
- Harris, A., Tran, T., Jenkins, S., Su, A., He, L., Zhu, Y. and Gane, S. (2022), "An appropriate verbal probability lexicon for communicating surgical risks is unlikely to exist", Journal of Experimental Psychology: Applied, 28(3): 538-554.
- Hausman, J. (2012), "Contingent valuation: from dubious to hopeless", *Journal of Economic Perspectives*, 26(4): 43–56.
- Hearld L.R., Alexander J.A., Shi Y., C. L. P. (2014). Pay-for-performance and public reporting program participation and administrative challenges among small- and medium-sized physician practices. *Medical CareResearch and Review*, *71*(3), 299–312. https://doi.org/10.1177/1077558713509018
- Herber, O. R., Gies, V., Schwappach, D., Thürmann, P., & Wilm, S. (2014). Patient information leaflets: informing or frightening? A focus group study exploring patients' emotional reactions and subsequent behavior towards package leaflets of commonly prescribed medications in family practices. *BMC Family Practice*, *15*(1), 163.
- Hirshleifer, J., & Riley, J. G. (1979). The analytics of uncertainty and information—an expository survey. *Journal of Economic Literature*, *17*(4), 1375-1421.
- Hocking J.S., Temple-Smith M., van Daniel M., Law M., Guy R.,Bulfone L.,Wood A., Low N., Donovan B., Fairley C.K., Kaldor J., Gunn J. (2016). Can preventive care activities in general practice be sustained when financial incentives and external audit plus feedback are removed? A cluster randomized controlled trial protocol. *Implementation Science*, *11*(1), Article number 122. DOI: https://doi.org/10.1186/s13012-016-0489-0

- Hollander M.J. (2015). Incentive Based Primary Care: Cost and Utilization Analysis. *The Permanente Journal*, *19*(4), 46–56. DOI: https://doi.org/10.7812/TPP/15-045
- Honda H, Kagawa R, Shirasuna M.(2022) On the round number bias and wisdom of crowds in different response formats for numerical estimation. *Sci Rep*. 2022;12:8167. DOI: https://doi.org/10.1038/s41598-022-11900-7
- Houle S.K.D., McAlister F.A., Jackevicius C.A, et al(2012). Does performance-based remuneration for individual health care practitioners affect patient care? A systematic review. Ann InternMed. 2012;157(12):889–99. DOI : 10.7326/0003-4819-157-12-201212180-00009.
- Howdon D, Rice N. Health care expenditures, age, proximity to death and morbidity: implications for an ageing population. *J Health Econ*. 2018;**57**:60-74.
- Husereau D, Boucher M, Noorani H. Priority setting for health technology assessment at CADTH. *Int J Technol Assess Health Care*. 2010;26:341-7.
- Hwang, C. L., & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag.
- Iskrov G, Stefanov R. Criteria for Drug Reimbursement Decision-Making: An Emerging Public Health Challenge in Bulgaria. *Balkan Med J*. 2016 Jan;33(1):27-35.
- J.R. Repullo Labrador y J.M. Freire Campo (2024). Performance evaluation and compensation in directly managed health centres. Part 2: The National Health System. SESPAS Report 2024, Gaceta Sanitaria. DOI: https://doi.org/10.1016/j.gaceta.2024.102368
- Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare Informatics Research*, *19*(2), 79–85.
- Johnson R.M.,Johnson T.,Zimmerman S.D.et al (2015). Outcomes of a seven practice pilot in a pay for performance (P4P)-based program in Pennsylvania. *Journal of Racial and Ethnic Health Disparities*, *2*(1), 139–48. https://doi.org/10.1007/s40615-014-0057-8

- Johnson, M., Cross, L., Sandison, N. (2018). Funding and policy incentives to encourage implementation of point-of-care C-reactive protein testing for lower respiratory tract infection in NHS primary care: A mixed-methods evaluation. *BMJ Open*, *8*(10), e024558. DOI: https://doi.org/10.1136/bmjopen-2018-024558
- Kahneman, D., Knetsch, J.L. (1992), "Valuing public goods: The purchase of moral satisfaction", *Journal of Environmental Economics and Management*, 2(1): 57-70.
- Kecmanovic, Milica;P.Hall, J. (2015). The use of financial incentives in Australian general practice. *MJA*, *202*(9), 488–92. DOI: https://doi.org/10.5694/mja14.00252
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press DOI: https://doi.org/10.1017/CBO9781139174084
- Kirschner K., Braspenning J., Jacobs J.E (2012) Design choices made by target users for a pay for performance program in primary care: an action research approach. BMC FamPract. 2012; 13 (article number 25).
- Kirschner K.,Braspenning J.,Akkermans R.P.Jacobs J.E. Annelies, Grol R.(2013). Assessment of pay for performance program in primary care designed by target users. *Family Practice*, *30*(2), 161–71. DOI: https://doi.org/https://doi.org/10.1093/fampra/cms055
- Kirschner K.,Braspenning J.,Jacobs J.E.A, Akkermans R. P., Grol R..(2013). Experiences of general practices with a participatory pay for performance program: A qualitative study in primary care. Aust J Prim Health. 2013;19 (2):102–6.
- Klarman, H. E., Francis, J. O., & Rosenthal, G. D. (1968). Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. *Medical Care*, *6*(1), 48-54.
- Klein, R. (2013). *The New Politics of the NHS: From Creation to Reinvention* (7th ed.). Radcliffe Publishing.
- Kling, C.L., D.J. Phaneuf and J. Zhao (2012), "From Exxon to BP: Has Some Number Become Better Than No Number?", *Journal of Economic Perspectives*, 26(4): 3-26. Latvala, T. and Kola, J. (2000), "Consumers' Willingness to Pay for Information about

Food Safety and Quality: Case Beef", Contributed paper at the International Food and Agribusiness Management Association IAMA 10 th World Congress and Forum, Chicago, USA. (http://agecon.tamu.edu/iama/2000Congress/2000_forum_papers.htm).

- Kolozsvari L.R.,Orozco-Beltran D., R. I. (2014). Do family physicians need more payment for working better? Financial incentives in primary care. *Atencion Primaria*, *46*(5), 261–66. DOI: https://doi.org/https://doi.org/10.1016/j.aprim.2013.12.014
- Kolozsvari L.R.,Rurik, I. (2013). Quality improvement in primary care.Financial incentives related to quality indicators in Europe. *OrvosiHetilap*, *154*(28), 1096–1101. DOI: https://doi.org/10.1556/OH.2013.29631
- Kontopanelis E.,Springate D.A.,Ashcroft D.M.,Valderas J.M., Sabine N. Van der Veer, Reeves D., Guthrie B.,Doran T. (2016). Associations between exemption and survival outcomes in the UK´s primary a care pay for performance programme: A restrospective cohort study. *BMJ Quality and Safety*, *25*(9), 657–70. https://doi.org/10.1136/bmjqs-2015-004602
- Krauth C.,LierschS.,Jensen S., A. V. . (2016). Would german physicians opt for pay for performance programs? A willingness-to-accept experiment in a large general practitioner´s sample. *Health Policy*, *120*(2), 148–158. DOI: https://doi.org/https://doi.org/10.1016/j.healthpol.2016.01.009
- Kuhnle, S., & Sander, A. (2010). The Emergence of the Western Welfare State. In F. G. Castles, S. Leibfried, J. Lewis, H. Obinger, & C. Pierson (Eds.), *The Oxford Handbook of the Welfare State*. Oxford University Press.
- LaaRocca A., Hoholm T. (2017). Coordination between primary and secondary care: the role of electronic messages and economic incentives. *BMC Health Services Research*, *17*(1), 149.
- Lauenroth VD, Stargardt T (2017). Pharmaceutical pricing in Germany: how is value determined within the scope of AMNOG? *Value Health*, 20:927-35.
- Lavergne MR, Law MR, Peterson S, Garrison S., Hurley J.,Cheng L., McGrail K.(2018).Effect of incentive payments on chronic disease management and health

services use in British Columbia, Canada: Interrupted time series analysis.HealthPolicy. 2018 Feb; 122(2):157-64.DOI: https://doi.org/10.1016/j.healthpol.2017.11.001

- Lavergne, M. R., Law M. R., Peterson S., Garrison S , Hurley J., Cheng L., McGrail K (2016).A population-based analysis of incentive payments to primary care physicians for the care of patients with complex disease.*CMAJ*. 2016;*188*(15), E375–83. http://doi.org/10.1503/cmaj.150858
- Lehtovuori T., Kauppila T.,Kallio J.,Raina M.,Suominen L. HA. Financial team incentives improved recording of diagnoses in primarya care: a quasi-experimental longitudinal follow-up study with controls. BMC Res Notes. 2015; 8:668.
- Lemak, C.H., Cohen, G.R., & Erb, N.D. (2013). Engaging Primary Care Physicians in Quality Improvement: Lessons From a Payer-Provider Partnership. *Journal of Healthcare Management, 58*, 429-443.
- Leshem, D. (2013). Oikonomia Redefined. *Journal of the History of Economic Thought, 35*(1), 43-61.
- Lester H.,MatharuT.,Mohamed M.A et al (2013). Implementation of pay for performance in primary care: A qualitative study 8 years after introduction. *British Journal of General Practice*, *63*(611), e408–15. DOI: https://dx.doi.org/10.3399%2Fbjgp13X668203
- Lin Y, Yin S, Huang J. Impact of pay for performance on behaviour of primary care physicians and patient outcomes. J Evid [Internet]. 2016 [consulted on 21/12/2018];9(1):8–23. Available in:: https://onlinelibrary.wiley.com/doi/abs/10.1111/jebm.12185
- Loewenstein, G., Brennan, T., & Volpp, K. G. (2007). Asymmetric Paternalism to Improve Health Behaviors. *Journal of the American Medical Association, 298*(20), 2415-2417.
- MacCrimmon, K., Smith, M. (1986), "Imprecise equivalences: Preference reversals in money and probability". Working paper, University of British Columbia, Vancouver, BC, Canada.

- Mahmood K (2003). The implementation of a local primary care incentive scheme: Medicalvs management priorities. Clin Gov An Int J. 2003;8(1):19–25.
- Marcotte L.,Hodlofski -A.,Bond A., Pankaj P., Sacks L., Navathe A.S.(2017). Into practice: How Advocate Health System uses behavioural economics to motivate physicians in its incentive program. *Healthc (Amst)*, *5*(3), 129–35. DOI: https://doi.org/http://dx.doi.org/10.1016/j.hjdsi.2016.04.011
- Marmot, M. (2005). Social Determinants of Health Inequalities. *The Lancet*, *365*(9464), 1099-1104.
- Márquez-Peláez S, Espín Balbino J, Olry de Labry Lima A, Benítez Hidalgo V, y grupo de trabajo RedETS paran MCDA. Guía para la elaboración de recomendaciones basada en análisis de decisión multicriterio. Sevilla: AETSA, Evaluación de Tecnologías Sanitarias de Andalucía. Madrid: Red Española de Agencias de Evaluación de Tecnologías Sanitarias; 2020.
- Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decisión analysis. *J Public Health*. 2013;35:460-6.
- Marsh K, Goetghebeur M, Thokala P, Baltussen R. *Multi-criteria decision analysis to support healthcare decisions*. New York: Springer; 2017.
- Marsh K, Ijzerman M, Thokala P, Baltussen R, Boysen M, Kaló Z, et al. Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19:125-37.
- Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. *PharmacoEconomics*. 2014;32:345-65.
- Martens J.D., Werkhoven M.J., Severens J.L (2007). Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. J. Eval. Clin. Pract. 2007;13(3):369–73.

- McBride-Stewart S.P.,Elton R (2008). Do quality incentives change prescribing patterns in primary care? An observational study in Scotland. Fam Pract. 2008;25(1):27–32.
- McDonald R. RM (2009). Pay for performance in primary care in England and California: Comparison of unintended consequences. Ann Fam Med. 2009;7(2):121–7.
- McDonald R.,Harrison S. Checkland K (2008). Incentives and control in primary healthcare: Findings from English pay for performance case studies. J Health Organ Manag. 2008; 22(1):48–62.
- McDonald R.,Harrison S.,Checkland K., Campbell S.M. (2007). Impact of financial incentives on clinical autonomy and internal motivation in primary care: Ethnographic study. Br Med J. 2007; 334(7608):1357–9.
- McKee, M., & Stuckler, D. (2020). If the World Fails to Protect the Economy, COVID-19 Will Damage Health Not Just Now But Also in the Future. *Nature Medicine*, *26*(5), 640-642.
- Ministry of Health, Consumer Affairs and Social Welfare .Studies and reports on health spending. [Internet]; [consultado el 15/11/2019]. Available in : Ministerio de Sanidad - Sanidad en datos - Estadisticas y Estudios
- Mirelman A, Mentzakis E, Kinter E, Paolucci F, Fordham R, Ozawa S, et al. Decision-Making Criteria among National Policy makers in Five Countries: A Discrete Choice Experiment Eliciting Relative Preferences for Equity and Efficiency. *Value Health*. 2012;15:534-9.
- Mitchell, R.C. and Carson, R.T. (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*, *Resources for the Future*. New York, NY: Resources for the Future.
- Moreno V.M.,Cermeño P.C., Gadea J.E. (2006). Incentives for primary care family doctors. Reasons for change and proposal of a new model. *Revista de Administración Sanitaria*, *4*(1), 119–53. DOI: //doi.org/10.1001/journalofethics.2015.17.7.stas1-1507
- Moreno, L. (2001). La "vía media" española del modelo de bienestar mediterráneo. *Papers: revista de sociología*, *63/64*, 67-82.

- Neil A.L., Nelson M., P. A. J. (2016). The new Australian after-hours general practice incentive payment mechanism: equity for rural general practice. *HealthPolicy*, *120*(7), 809–17. https://doi.org/10.1016/j.healthpol.2016.05.005
- Neil A.L., Nelson M.R.,Richardson T.,Mann-Leonard M., Palmer A.J. (2015). General practice after-hours incentive funding: A rationale for change. *Medical Journal of Australia*, *203*(2), 82–6. https://doi.org/10.5694/mja14.01229
- Newhouse, J. P. (1993). *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, MA: Harvard University Press.
- NHS Quality-Outcome Framework; Report of July(2018).[consulted on 05/09/2019]; pág 29. Disponible en: https://www.england.nhs.uk/wp-content/uploads/2018/07/quality-outcome-framework-report-of-the-review.pdf
- NICE (National Institute for Health and Care Excellence). Guide to the methods of technology appraisal 2013. Available from: https://www.nice.org.uk/process/pmg9
- Norman A.H.,Russell A.J., M. J. (2014). The payment for performance model and its influence on british general practitioners principles and practice. *Cadernos de Saude Publica*, *30*(1), 55–67. DOI: https://doi.org/10.1590/0102-311x00149912
- OECD. *New Health Technologies: Managing Access, Value and Sustainability*. Paris: OECD Publishing; 2011.
- Okunade AA, Murthy VNR. Technology as a 'major driver' of health care costs: a cointegration analysis of the Newhouse conjecture. *J Health Econ*. 2002;21:147-59.
- Oliveira Martins J, de la Maisonneuve C. The Drivers of Public Expenditure on Health and Long-Term Care: An Integrated Approach. *OECD Economic Studies* No 43, 2006/2.
- Olivier Saint-Lary, Claire Leroux, Cécile Dubourdieu, Cécile Fournier and Irène François-Purssell (2015). Patients' views on pay for performance in France: a qualitative study in primary care. British Journal of General Practice 2015; 65 (637): e552-e559. DOI: https://doi.org/10.3399/bjgp15X686149

- Oortwijn W, Sampietro-Colom L, Habens F. Developments in value frameworks to inform the allocation of healthcare resources. *Int J Technol Assess Health Care*. 2017;33:323-9.
- Park, S., & Han, E. (2016). Do physicians change prescription practice in response to financial incentives? *International Journal of Health Services*, *46*(3), 531–46. https://doi.org/10.1177/0020731416649846
- Parkin E. (2018). General Practice in England. Briefing Paper nº. 07194, 18 September 2018. House of Commons Library.
- Payne G, Laporte A, Deber R, Coyte PC.(2007) Counting backward to health care's future: using time-to-death modelling to identify changes in end-of-life morbidity and the impact of aging on health care expenditures. *Milbank Q*. 2007;85:213-57.
- Pecchia L, Martin JL, Ragozzino A, Vanzanella C, Scognamiglio A, Mirarchi L, Morgan S.P.(2013) . User needs elicitation via analytic hierarchy processs (AHP). A case study on a Computed Tomography (CT) scanner. *BMC Med Inform Decis Mak*. 2013;13:2. DOI: https://doi.org/10.1186/1472-6947-13-2
- Peckham S., W. A. (2010). Pay for performance schemes in primary care: what have we learnt? *Quality in Primary Care*, *18*(2), 111–116.
- Phillips LD, Fasolo B, Zafiropoulos N, Beyer A. Is quantitative benefit risk modeling of drugs desirable or possible? *Drug Discovery Today Technol*. 2011;8:e3–10.
- Pierson, P. (2001). *The New Politics of the Welfare State*. Oxford University Press.
- Pinto, J. L., Sánchez, F. I., & Rovira, J. (1998). Medición de los beneficios intangibles de un tratamiento médico a través de la valoración contingente. *Hacienda Pública Española*, *147*, 51-62.
- Pinto-Prades, J.-L., Martínez-Pérez, J.-E. and Abellán-Perpiñán, J.-M. (2006), "The influence of the ratio bias phenomenon on the elicitation of health states utilities", *Judgment and Decision Making*, 1(2): 118–133.

- Pinto-Prades, J.L., Sánchez-Martínez, F.I., Abellán-Perpiñán, J.M. and Martínez-Pérez, J.E. (2018), "Reducing preference reversals: The role of preference imprecision and nontransparent methods", *Health Economics*, 27(8): 1230–46.
- Poe, G.L. (2016), "Behavioral Anomalies in Contingent Values and Actual Choices", *Agricultural and Resource Economics Review*, 45(2): 246–269.
- Polain M, Franken M, Koopmanschap M, Cleemput I. *Drug reimbursement systems: international comparison and policy recommendations*. Brussels: Belgian Health Care Knowledge Centre (KCE); 2010.
- Porter ME. What is Value in Health Care. *N Engl J Med.* 2010; 363:2477-2481.
- Protiere, C., Donaldson, C., Luchini, S., Moatti, J.P. and Shackley, P. (2004), "The impact of information on non-health attributes on willingness to pay for multiple health care programs", *Social Science & Medicine*, 58: 1257-1269.
- Quereshi N.,Weng S., H. N. (2016). The role of cost-effectiveness analysis in the development of indicators to support incentive-based behaviour in primary care in England. *Journal of Health Services Research and Policy*, *21*(4), 263–271. DOI: https://doi.org/10.1177/1355819616650912
- Rezaei J, Arab A, Mehregan M. Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research. *J Behav Decis Mak*. 2022;35:e2262.
- Robinson J. (2001) Theory and practice in the design of physician payment incentives. *Milbank Quarterly III*, 79(2): 149–177. DOI: https://doi.org/10.1111/1468-0009.00202
- Roland M, Elliott M, Lyratzopoulos G, Barbiere J, Parker R A, Smith P et al (2009). Reliability of patient responses in pay for performance schemes: analysis of national General Practitioner Patient Survey data in England *BMJ* 2009; 339: b3851 DOI: https://doi.org/10.1136/bmj.b3851 (Published 29 September 2009)
- Roland, M., Campbell, SM., Bailey, N., Whalley, D., & Sibbald, BS. (2006). Financial incentives to improve the quality of primary care in the consequences of change. *Primary Health Care Research and Development , 1*.

- Roy, B. (1968). Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *Revue française d'informatique et de recherche opérationnelle, 2*(8), 57-75.
- Russel T.,Petterson S., Klink K. (2015). Primary care physicians are more likely to participate in medicare incentives than other eligible physicians. *American Family Physician, 92*(3), 182.
- Rutten-Van Mölken M, Leijten F, Hoedemakers M, Tsiachristas A, Verbeek N, Karimi M,Bal R.,Antoinette de Bont, Islam K., Askildsen J.E., Czypionka T., Kraus M., Huic M., Pitter J.G., Vogt V.,Stokes J., Baltaxe E., Selfie Consortium. (2018) .Strengthening the evidence-vase of integrated care for people with multi-morbidity in Europe using Multi-Criteria Decision Analysis (MCDA). *BMC Health Serv Res*. 2018;18:576.DOI: https://doi.org/10.1186/s12913-018-3367-4
- Ryan A.M.,Krinsky S., Kontopanelis E. (2016). Long-term evidence for the effect of pay for performance in primary care on mortality in the UK: a population study. *The Lancet, 388*(10041), 268–74. https:/doi.org/10.1016/S0140-6736(16)00276-2
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill.
- Saint-Lary O.,Bernard E.,Sicsic J., Plu I., Francois-Pursell I., Franc C. (2013). Why did most french GP choose not to join the voluntary national pay-for-performance program? *PLoS ONE, 8*(9), Article number e72684. DOI: https://doi.org/https://doi.org/10.1371/journal.pone.0072684
- Saint-Lary O.,Plu I., N. M. (2012). Ethical Issues raised by the introduction of payment for performance in France. *Journal of Medical Ethics, 38*(8), 485–91. https://doi.org/https://dx.doi.org/10.1136%2Fmedethics-2011-100159
- Saint-Lary, O., Leroux, C., Dubourdieu, C., Fournier, C., & François-Purssell, I. (2015). Patients' views on pay for performance in France: a qualitative study in primary care. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners, 65*(637), e552–e559.

- Sánchez-Martínez, F.I., Martínez-Pérez, J.E., Abellán-Perpiñán, J.M., Pinto-Prades, J.L. (2021), "The value of statistical life in the context of road safety: new evidence on the contingent valuation/standard gamble chained approach", *Journal of Risk and Uncertainty*, 63(2): 203-228.
- Schmets G, Rajan D, Kadandale S, editors. *Strategizing national health in the 21st century: a handbook*. Geneva: World Health Organization; 2016.
- Schwartz, L. M., Woloshin, S., & Welch, H. G. (2009). Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Annals of Internal Medicine*, *150*(8), 516-527.
- Scott A, Sivey P, Ait Ouakrim D, Willemberg L.,Nacarella L., Furler J., Young D. (2011). The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database of Systematic Reviews. 2011; Issue 9. Art. No.: CD008451. DOI: 10.1002/14651858.CD008451.pub2
- Scott A,Schurer S, Jensen PH, Sivey P. (2009) The effects of an incentive program on quality of care in diabetes management, Health Economics, 18(9): 1091–1108.
- Scott, S. D., Hirschinger, L. E., Cox, K. R., McCoig, M., Brandt, J., & Hall, L. W. (2009). The natural history of recovery for the healthcare provider "second victim" after adverse patient events. *Quality & Safety in Health Care*, *18*(5), 325–330.
- Shah S.M.,Carey I.M., Harris T.,Dewilde S.,Cook D.G. (2011). Quality of chronic disease care for older people in care homes and the community in a primary care pay for performance system: retrospective study. *BMJ, 342:7797*(587), Article number d912. https://doi.org/10.1136/bmj.d912
- Shane D.,MacKinney A.C.,Ullrich F.,Mueller K.J., Weigel P., Rural Health Research&Plicy Centers, RUPRI Center for Rural Health Policy Analysis, University of Iowa College of Public Health, Department of Health Management and Policy (2013). Assessing the impact of rural provider services mix on the primary care incentive payment program. *Rural PolicyBrief, (2013 16)*, 1–6.

- Shono, A., Kondo, M., Ohmae, H. and Okubo, I. (2014), "Willingness to Pay for Public Health Services in Rural Central Java, Indonesia: Methodological Considerations When Using the Contingent Valuation Method", *Social Science & Medicine*, 110: 31–40.
- Sirota, M. and Juanchich, M. (2019), "Ratio format shapes health decisions: the pratical significance of the 1 in X effect", *Medical Decision Making*, 32(2): 179-193.
- Sirota, M., Juanchich, M., Petrova, D., García-Retamero, R., Walasek, L. and Bhatia, S. (2018), "Health Professionals Prefer to Communicate Risk-Related Numerical Information using ''1-in-X'' Ratios", *Medical Decision Making*, 38(3): 1-11.
- Smith S, Newhouse J, Freeland M. Income, Insurance, and Technology: Why Does Health Spending Outpace Economic Growth? *Health Aff*. 2009;28:1276-84.
- Steel N., Maisey S., Clark A., Fleetcroft R., Howe A., (2007). *Quality of clinical primary care and targeted incentive payments: An observational study*. *57*(539), 449–54.
- Stewart, R. E., Marcus, S. C., Hadley, T. R., Hepburn BM., Mandell D.S.(2018). State adoption of incentives to promote evidence-based practices in behavioural health systems. *Psychiatric Services*, *69*(6), 685–88. DOI: https://doi.org/10.1176/appi.ps.201700508
- Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. *Multi-criteria decision analysis to value orphan medicines*. Office of Health Economics. 2013; Research paper 13/03.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.
- Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, Longrenn T, Mussen F, Peacock S, Watkins J, Ijzerman M.(2016) Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19:1-13.
- Tim Doran, Catherine Fullwood, Hugh Gravelle, David Reeves, Evangelos Kontopanelis, Urara Hiroeh, Martind Roland. (2006). Pay-for-performance programs in

family practices in the United Kingdom. *New England Journal of Medicine*, *355*(4), 375–84. https://doi.org/10.1056/NEJMsa055505

- 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.
- Toumi M, Remuzat C, El Hammi E, Millier A, Aballéa S, Chouaid C, et al. Current process and future path for health economic assessment of pharmaceuticals in France. *J Mark Access Health Pol*. 2015;3:27902.
- Tversky, A., Kahneman, D. (1974), "Judgment under uncertainty: Heuristics and biases", *Science,* 185: 1124–1130.
- Tversky, A., Sattath, S., Slovic, P. (1988), "Contingent weighting in judgment and choice", *Psychological Review*, 95(3): 371–384.
- Urbanos-Garrido, R. (2016). La desigualdad en el acceso a las prestaciones sanitarias. Propuestas para lograr la equidad [Inequality in access to health care services. Policy recommendations aimed at achieving equity]. *Gaceta Sanitaria, 30* Suppl 1, 25–30.
- van Til JA, Renzenbrink GJ, Dolan JG, Ijzerman MJ. The use of the analytic hierarchy process to aid decisión making in acquired equinovarus deformity. *Arch Phys Med Rehabil*. 2008;89:457–62.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- Wahlster P, Goetghebeur M, Kriza C, Niederländer C, Kolominsky-Rabas P and on behalf of the National Leading-Edge Cluster Medical Technologies 'Medical Valley EMN' Balancing costs and benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA). *BMC Health Serv Res*. 2015;15:262.
- Webster, R., Weinman, J., Rubin, G. (2018), "Positively Framed Risk Information in Patient Information Leaflets Reduces Side Effect Reporting: A Double-Blind Randomized Controlled Trial", *Annals of Behavioral Medicine*, 52: 920-929

- Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., & Luce, B. R. (2003). Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health, 6*(1), 9-17.
- Wettermark B., Pehrsson A., Juhasz-Haverinen M., Veg A.,Edlert M., Törnwall-Bergendahl G., Almkvist H., Godman B., Granath F., Bergman U. (2009). Financial incentives linked to self-assessment of prescribing patterns: A new approach for quality improvement of drug prescribing in primary care. Qual Prim Care. 2009; 17(3):179–89.
- Willemé P, Dumont M. Machines that go 'ping': medical technology and health expenditures in the OECD countries. *Health Econ*. 2014;24:1027-41; Corrected and republished from: *Health Econ*. 2015;24:387-8.
- Williams, A. (1985). Economics of coronary artery bypass grafting. *British Medical Journal 291*(6491), 326-329.
- Wolfe A. Institute of Medicine Report: Crossing the Quality Chasm: A New Health Care System for the 21st Century. Policy, Politics, & Nursing Practice. 2001;2(3):233-235. doi:10.1177/152715440100200312
- World Bank. (1993). *World Development Report 1993: Investing in Health*. New York: Oxford University Press.
- Wright M. Pay for performance programs: Do they improve the quality of primary care? AustFam Physician. 2012;41(12):989–91.
- Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Applied Cognitive Psychology, 11*, 495-506.
- Yi, Y., Stvilia, B., and Mon, L. (2012), "Cultural influences on seeking quality health information: An exploratory study of the korean community", *Libray and Information Science Research*, 34(1): 45-51.
- Youngkong, S., Baltussen, R., Tantivess, S., Mohara, A., & Teerawattananon, Y. (2012). Multicriteria decision analysis for including health interventions in the universal health coverage benefit package in Thailand. *Value in Health, 15*(6), 961-970.

- Zeckhauser, R. & Shepard, D. (1976). Where Now for Saving Lives?. *Law and Contemporary Problems*, 40, 5-45
- Zelei T, Mendola ND, Elezbawy B, Némeh B, Campbell JD. Criteria and Scoring Functions Used in Multi-criteria Decision Analysis and Value Frameworks for the Assessment of Rare Disease Therapies: A Systematic Literature Review. *Pharmacoeconomics open*. 2021;5:605-12.
- Zhang X., S. A. (2018). Blended capitation and incentives: Fee codes inside and outside the capitated basket. *Journal of Health Economics*, *60*, 16–29. http://doi.org/10.1016/j.jhealeco.2018.03.002
- Zozaya N. Arrizubieta MI, Bollo E, Castellví I, Espín J, Ortego N, et al. A multi-criteria decision analysis on the value of nintedanib for interstitial lung diseases. *Int J Technol Assess Health Care*. 2022;38:e64.
- Zweifel P, Steinmann L, Eugster P. The Sisyphus syndro e in health revisited. Int J Health Care Finance Econ. 2005;5:127-45.