



Universitat de Lleida

Models matemàtics i estadístics aplicats a la presa de decisions en salut

Carles Forné Izquierdo

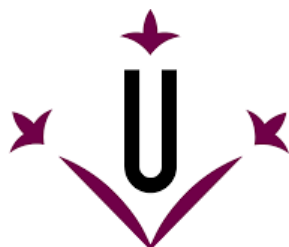
<http://hdl.handle.net/10803/670064>



Models matemàtics i estadístics aplicats a la presa de decisions en salut està subjecte a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 4.0 No adaptada de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Les publicacions incloses en la tesi no estan subjectes a aquesta llicència i es mantenen sota les condicions originals.

(c) 2020, Carles Forné Izquierdo



UNIVERSITAT DE LLEIDA

TESI DOCTORAL

**Models matemàtics i estadístics
aplicats a la presa de decisions
en salut**

Autor:
Carles FORNÉ
IZQUIERDO

Directora i tutora:
Dra. Montserrat RUÉ
MONNÉ

*Memòria presentada per optar al grau de doctor per la Universitat de
Lleida dins el programa de Doctorat en Salut*

9 de juny de 2020

Declaració d'autoria

Jo, Carles FORNÉ IZQUIERDO, declaro que aquesta tesi, titulada «Models matemàtics i estadístics aplicats a la presa de decisions en salut», i el treball que s'hi presenta són els meus. Confirmo que:

- Aquest treball es va realitzar totalment o principalment durant la meva candidatura al títol de doctor a la Universitat de Lleida.
- Allà on he hagut de consultar treballs publicats per altres, aquests sempre estan clarament atribuïts.
- On he citat el treball dels altres, sempre se'n dona la font. Amb l'excepció d'aquestes citacions, aquesta tesi és totalment el meu propi treball.
- On la tesi es basa en el treball realitzat per mi mateix conjuntament amb altres persones, he posat de manifest exactament el que van fer els altres i el que hi he aportat jo mateix.

Signat:

Data:

«To consult the statistician after an experiment is finished is often merely to ask him to conduct a *post mortem* examination. He can perhaps say what the experiment died of.»

Sir Ronald Fisher

Presidential Address to the First Indian Statistical Congress, 1938

«Lo bueno, si breve, dos veces bueno; y aun lo malo, si poco, no tan malo.»

Baltasar Gracián a l'*Oráculo manual y arte de prudencia*

Osca: Juan Nogués, 1647

UNIVERSITAT DE LLEIDA

Resum

Facultat de Medicina

Departament de Ciències Mèdiques Bàsiques

Doctor

Models matemàtics i estadístics aplicats a la presa de decisions en salut

per Carles FORNÉ IZQUIERDO

La presa de decisions en salut ha de basar-se en la millor evidència rellevant disponible, més si aquestes decisions incideixen en tota la societat mitjançant l'aplicació de polítiques públiques de salut i, donat que totes les persones són pacients potencials — presents o futurs —, també en les decisions clíniques. Prendre la decisió òptima depèn de molts factors que han de ser degudament identificats i minuciosament mesurats en les diferents alternatives d'elecció, entre els quals, l'escassetat de recursos disponibles per a cobrir una il·limitada demanda de salut. L'avaluació econòmica en salut té per objecte informar i facilitar decisions sobre quin és el millor ús dels recursos limitats disponibles, amb la finalitat de maximitzar-ne els beneficis en salut, prioritizant les alternatives d'elecció mitjançant criteris d'eficiència. Actualment, les agències nacionals d'avaluació de tecnologies sanitàries basen les seves recomanacions de finançament en estudis d'avaluació econòmica fonamentats en la millor evidència disponible i metodològicament correctes. Això inclou des d'un disseny adequat a l'objectiu, fins a un report complet i transparent que permeti la reproductibilitat dels resultats, passant pel rigorós tractament analític de les dades d'estudi. En aquest marc, els models analítics són una font d'evidència vàlida per a la presa de decisions, com ho són els assajos clínics i els estudis observacionals, també ben dissenyats i ben realitzats. En aquesta tesi s'inclouen quatre treballs en què l'aplicació dels mètodes analítics adequats van permetre donar resposta a uns objectius que d'altra manera no podrien resoldre's, o només parcialment o amb majors limitacions.

Gràcies a l'ús de models probabilístics es va poder avaluar el cost-efectivitat i el dany-benefici de 2.625 escenaris diferents de cribratge de càncer de mama, i es mostrà que les estratègies basades en el risc poden reduir els danys i els costos en comparació amb les estratègies uniformes. Donat que per avançar en l'organització de programes de cribratge basats en el risc cal desenvolupar mesures precises de risc individual de càncer de mama, es va formular un model conjunt bayesià que combina la història longitudinal de la densitat mamària i l'edat en el diagnòstic de càncer de mama. La metodologia

proposada podria aplicar-se per obtenir mesures més precises del risc individual de càncer de mama. Per això caldria realitzar estudis de cohorts de base poblacional més grans, amb disposició de mesures longitudinals de la densitat mamària, així com d'altres factors de risc coneguts, com les lesions de mama benignes i el risc poligènic. D'aquesta manera, seria possible assolir capacitats predictives aplicables a la personalització del cribratge de càncer de mama, o per establir estratègies de prevenció per a dones amb risc elevat.

En aquesta tesi també s'inclou una anàlisi de dinou potencials biomarcadors cardiovasculars en població amb malaltia renal mitjançant l'aplicació d'algorismes d'aprenentatge automàtic. En concret, aplicant l'algorisme de bosc aleatori de supervivència per a esdeveniments competitiu, va ser possible avaluar la capacitat predictiva de tots els biomarcadors alhora, junt amb tota la resta de factors de risc coneguts, altrament impossible donat el limitat nombre d'esdeveniments cardiovasculars disponibles per ajustar els habituals models de supervivència multivariables. Si bé l'addició dels biomarcadors potencials va millorar marginalment la capacitat de discriminació obtinguda amb només els factors de risc cardiovascular coneguts, es van identificar els biomarcadors més prometedors, als quals idealment haurien de ser destinats els recursos de la futura recerca, sempre que el benefici potencial de la nova evidència superi els costos de la investigació.

Per acabar, s'inclou l'avaluació de l'impacte en costos i en salut de l'augment de l'ús de la intervenció coronària percutània en pacients ingressats amb síndrome coronària aguda mitjançant un model analític semi-Markov, ja que els registres poblacionals mostren una infrautilització de la intervenció en ancians, possiblement deguda a la manca d'estimacions de la seva eficiència o del seu cost-utilitat. El model desenvolupat conceptualitza el curs de la malaltia en un horitzó temporal de 8 anys des de l'ingrés hospitalari, i proporciona resultats molt robustos que indiquen que augmentar el nombre de pacients ancians que es beneficiïn de la intervenció és rendible en un llinar de disposició a pagar molt inferior als estàndards en països europeus. Per tant, el cost-utilitat no hauria de ser un motiu per no incrementar l'ús de la intervenció en aquesta població.

Al llarg dels treballs inclosos en aquesta tesi han quedat palesos el paper i la utilitat de la modelització matemàtica i estadística, aplicada de manera adequada i rigorosa segons els objectius de cada estudi, per proporcionar evidència útil que faciliti la presa de decisions, la planificació i la gestió en salut.

UNIVERSITY OF LLEIDA

Abstract

Faculty of Medicine

Department of Basic Medical Sciences

Doctor

Mathematical and statistical models applied to health decision-making

by Carles FORNÉ IZQUIERDO

Health decision-making should be based on the best relevant evidence available, even more so when these decisions impact society through the application of public health policies, and given that all people are also potential—present or future—patients in clinical decisions. Optimal decision-making depends on many factors that must be properly identified and precisely measured for each alternative choice. These include the scarce resources available to meet unlimited healthcare demand. The economic evaluation in health aims to inform and facilitate decision-making on the best use of the limited resources available, in order to maximize health benefits, prioritizing the alternative choices following efficiency criteria. Currently, national agencies for health technology assessment base their recommendations about whether or not to fund particular interventions on economic evaluation studies that incorporate the best relevant evidence available and are methodologically correct. This ranges from a proper study design suited to the objective, to a complete and transparent report allowing reproducibility of results, undertaking a rigorous analytical handling of the study data. In this framework, analytical models are a valid source of evidence for decision-making, such as well-designed and well-conducted clinical trials and observational studies. Four works are included in this thesis, in which the application of appropriate analytical methods allowed us to provide answers to objectives that otherwise could not be solved, or only partially or with greater limitations.

Thanks to the use of probabilistic models, the cost-effectiveness and harm-benefit of 2,625 different scenarios of screening for breast cancer were assessed, showing that risk-based strategies can reduce harm and costs compared to uniform strategies. Given that precise measures of individual breast cancer risk must be developed in order to shift toward personalized screening strategies based on breast cancer risk, a Bayesian joint model that combines the longitudinal history of breast density and age at diagnosis of breast cancer was formulated. The proposed methodology could be applied to obtain more precise measures of the individual risk of breast cancer. To achieve

this objective, it would be necessary to conduct studies of larger population-based cohorts, with longitudinal measurements of breast density as well as other known risk factors, such as benign breast disease and polygenic risk. In this way, it would be possible to achieve a predictive ability that makes it possible to personalize screening for breast cancer, or to establish prevention strategies for women at high risk.

An analysis of 19 potential cardiovascular biomarkers in patients with renal disease by means of machine learning algorithms is also included in this thesis. Specifically, by applying a random survival forest for competing risks, it was possible to evaluate the predictive ability of all biomarkers at the same time, along with all other known risk factors, otherwise impossible given the limited number of cardiovascular events to fit classic multivariable survival models. Although the addition of potential biomarkers marginally improved the predictive ability obtained with only the known cardiovascular risk factors, the most promising biomarkers were identified, where resources for future research should ideally be allocated, whenever the benefit of new evidence exceeds the costs of research.

Finally, the evaluation of costs and health outcomes of increasing the use of percutaneous coronary intervention in patients admitted with acute coronary syndrome through a semi-Markov analytical model is included, since population-based registries have described lower use of the intervention in the elderly, possibly due to the lack of estimates of their efficiency or cost-utility. The model developed conceptualizes the course of the disease over an 8-year time horizon from hospital admission, and provides very robust results showing that increasing the number of elderly patients who benefit from the intervention is cost-effective on a threshold of willingness-to-pay far below the standards in European countries. Therefore, cost-utility should not be a reason not to increase the use of percutaneous coronary intervention in this population.

Throughout the work included in this thesis, the role and usefulness of mathematical and statistical modeling has been shown, properly and rigorously applied according to the objectives of each study, to provide useful evidence that facilitates decision-making, planning and healthcare management.

UNIVERSIDAD DE LLEIDA

Resumen

Facultad de Medicina

Departamento de Ciencias Médicas Básicas

Doctor

Modelos matemáticos y estadísticos aplicados a la toma de decisiones en salud

por Carles FORNÉ IZQUIERDO

La toma de decisiones en salud debe basarse en la mejor evidencia relevante disponible, más si estas decisiones inciden en toda la sociedad mediante la aplicación de políticas públicas de salud y, dado que todas las personas son pacientes potenciales —presentes o futuros—, también en las decisiones clínicas. Tomar la decisión óptima depende de muchos factores que deben ser debidamente identificados y minuciosamente medidos en las diferentes alternativas de elección, entre ellos, la escasez de recursos disponibles para cubrir una ilimitada demanda de salud. La evaluación económica en salud tiene por objeto informar y facilitar decisiones sobre cuál es el mejor uso de los recursos limitados disponibles, con el fin de maximizar los beneficios en salud, priorizando las alternativas de elección mediante criterios de eficiencia. Actualmente, las agencias nacionales de evaluación de tecnologías sanitarias basan sus recomendaciones de financiación en estudios de evaluación económica fundamentados en la mejor evidencia disponible y metodológicamente correctos. Esto incluye desde un diseño adecuado al objetivo, hasta una comunicación completa y transparente que permita la reproducibilidad de los resultados, pasando por el riguroso tratamiento analítico de los datos de estudio. En este marco, los modelos analíticos son una fuente de evidencia válida para la toma de decisiones, como lo son los ensayos clínicos y los estudios observacionales, también bien diseñados y bien realizados. En esta tesis se incluyen cuatro trabajos en que la aplicación de los métodos analíticos adecuados permitió dar respuesta a unos objetivos que de otra manera no podrían resolverse, o sólo parcialmente o con mayores limitaciones.

Gracias al uso de modelos probabilísticos se pudo evaluar el coste-efectividad y el daño-beneficio de 2.625 escenarios diferentes de cribado de cáncer de mama, y se mostró que las estrategias basadas en el riesgo pueden reducir el daño y los costes en comparación con las estrategias uniformes. Dado que para avanzar en la organización de programas de cribado basados en el riesgo es necesario desarrollar medidas precisas de riesgo individual de cáncer de mama, se formuló un modelo conjunto bayesiano que combina la historia

longitudinal de la densidad mamaria y la edad en el diagnóstico de cáncer de mama. La metodología propuesta podría aplicarse para obtener medidas más precisas del riesgo individual de cáncer de mama. Para ello sería necesario realizar estudios de cohortes de base poblacional más grandes, con disposición de medidas longitudinales de la densidad mamaria, así como de otros factores de riesgo conocidos, como las lesiones de mama benignas y el riesgo poligénico. De este modo sería posible alcanzar capacidades predictivas aplicables a la personalización del cribado de cáncer de mama, o para establecer estrategias de prevención para mujeres con riesgo elevado.

En esta tesis también se incluye un análisis de diecinueve potenciales biomarcadores cardiovasculares en población con enfermedad renal mediante la aplicación de algoritmos de aprendizaje automático. En concreto, aplicando el algoritmo de bosque aleatorio de supervivencia para eventos competitivos, fue posible evaluar la capacidad predictiva de todos los biomarcadores al mismo tiempo, junto con todo el resto de factores de riesgo conocidos, de otro modo imposible dado el limitado número de eventos cardiovasculares disponibles para ajustar los habituales modelos de supervivencia multivariados. Si bien la adición de los biomarcadores potenciales mejoró marginalmente la capacidad de discriminación obtenida con sólo los factores de riesgo cardiovascular conocidos, se identificaron los biomarcadores más prometedores, a los que idealmente deberían ser destinados los recursos de la futura investigación, siempre que el beneficio potencial de la nueva evidencia supere los costes de la investigación.

Para acabar, se incluye la evaluación del impacto en costes y en salud del aumento del uso de la intervención coronaria percutánea en pacientes ingresados con síndrome coronario agudo mediante un modelo analítico semi-Markov, ya que los registros poblacionales muestran una infrautilización de la intervención en ancianos, posiblemente debida a la falta de estimaciones de su eficiencia o de su coste-utilidad. El modelo desarrollado conceptualiza el curso de la enfermedad en un horizonte temporal de 8 años desde el ingreso hospitalario, y proporciona resultados muy robustos que indican que aumentar el número de pacientes ancianos que se benefician de la intervención es rentable en un umbral de disposición a pagar muy inferior a los estándares en países europeos. Por tanto, el coste-utilidad no debería ser un motivo para no incrementar el uso de la intervención en esta población.

A lo largo de los trabajos incluidos en esta tesis han quedado patentes el papel y la utilidad de la modelización matemática y estadística, aplicada de manera adecuada y rigurosa según los objetivos de cada estudio, para proporcionar evidencia útil que facilite la toma de decisiones, la planificación y la gestión en salud.

Agraïments

A les portes de dipositar la tesi, em passen pel cap moltes persones que hi han intervingut d'alguna manera: aquelles que ho han fet directament mitjançant col·laboracions científiques; i les que ho han fet en la meua vida, la qual cosa també es reflecteix en aquest treball, potser d'una manera més subtil, però molt evident als meus ulls.

La meua vida professional i personal va canviar el dia que la Montse Rué, la meua directora i a qui tinc en enorme estima, va confiar en mi per incorporar-me al seu grup de recerca. De tot cor, crec que probablement és el millor que m'ha passat en la meua vida adulta, per la qual cosa li estic enormement agraït, d'ençà d'aquell juliol de 2010, ara i sempre. Gràcies a ella i a les companyes del grup vaig adquirir la passió pròpia per la recerca, duta a terme sempre amb honestat, rigor i transparència. No puc deixar d'agrair l'acolliment que també de bon inici vaig trobar en la resta de membres del grup: la Montse, l'Ester, la Cory i en Roger, que sempre m'han donat suport i m'han esperonat a tirar endavant la tesi. «Gent bona, bona gent», com diria el bo d'en Roger, a qui estic especialment agraït per sempre comptar amb mi a les calçotades que organitza al seu tros.

En un dels moments segurament més crítics de la tesi, vaig tenir la sort de conèixer els amics “bayes-cians”: la Carmen, l'Anabel, l'Hèctor i en Danilo. Excel·lents matemàtics i estadístics, no és tant per això sinó per la seva vàlua com a persones i la seva generositat, el motiu pel qual els estic enormement agraït. Sense ells mai hauríem pogut donar resposta a una part important d'aquesta tesi, i probablement mai podré agrair-los-ho prou. A més a més, donada la meua especial predilecció per un bon àpat entre amics, també els he d'agrair fer-me descobrir l'arròs al forn, una recepta valenciana —no podia ser d'una altra manera— que, tot sigui dit, segons el meu punt de vista, hauria d'encapçalar tots els llibres de receptes d'arròs del món.

No tinc més que paraules d'agraïment també per als companys del grup de recerca liderat per la Lupe Gómez. Tant la Lupe, com en Carles, la Malu, en Klaus i tota la resta de membres van acollir-nos en el seu grup, i així hem tingut l'oportunitat i el plaer de compartir sessions estadístiques, projectes d'investigació i bons moments tots plegats.

El meu sincer agraïment a en José Manuel i la Irene Román per la confiança que han dipositat en mi donant-me l'oportunitat de col·laborar en els seus objectius de recerca en la recta final de la tesi.

Sense cap mena de dubte ha estat de gran transcendència per a mi el meu pas per la consultoria privada Oblikue. Les lliçons sobre economia, economia de la salut i modelització en avaluació econòmica que he rebut d'en Ramon i en Max, que compto com a bons amics, són pal de paller d'aquesta tesi. El meu pas per Oblikue també em va permetre créixer professionalment, i va suposar encara més satisfacció personal fer-ho amb grans companys i companyes que m'ho feien tot més fàcil i amb qui mantenim afecte mutu: la

Maribel, l'Alba, en Ferran, l'Estefany, en David, la Blanca... Segur que em deixo a algú, demano disculpes. A tots ells els he d'agrair la paciència que sempre han tingut amb mi.

En tot aquest trajecte m'han acompanyat diversos amics. He pogut gaudir, i espero seguir-ho fent, d'entretengudíssimes tertúlies estadístiques i no estadístiques amb els amics Albert i Jordi. Escodrinyar el paper de l'*Homo bioestadisticus* en això de la recerca biomèdica i la rigorositat en l'aplicació del mètode científic i les conclusions que se'n poden derivar ha estat una activitat recurrent per l'alt interès que ens despertaven aquests temes. D'aquestes tertúlies també n'he pogut extreure coneixement i experiència profitosos en el meu treball. Dels temes no estadístics és millor no deixar-ne document escrit. De la resta d'amics, tant de l'*old school* de la Facultat de Matemàtiques com dels amics que he tingut la sort de fer a Lleida, he de dir que, pel que sembla, finalment podré tancar aquesta etapa defensant la tesi malgrat l'aniquilació voluntària de neurones en incomptables sopars i festes. Gràcies amics!

No per ser els darrers dels meus agraïments són els menys importants. De fet, tot el contrari. Naturalment estic totalment en deute amb els meus pares, a qui més estimo, que em van donar la vida i l'etern amor incondicional. De la meua mare, a qui va dedicada aquesta tesi, espero haver après a fer front a la vida amb positivisme malgrat les adversitats, que el bon humor i l'amor és la millor via per ser feliç i fer feliç a qui estimes. Del meu pare, reconegut modelista ferroviari i fotògraf amateur des d'un taller mecànic del Poblenou, vaig poder aprendre com és de necessari estar ben atent als detalls per aconseguir un bon resultat. Dels meus estimats germans, en Joan i la Núria, vull destacar la fidelitat d'ell i la valentia i fermesa d'ella. També vull agrair al meu germà haver estat pare junt amb l'Inma, encomiable mare, perquè m'han fet tiet de la preciosa Noa, que ha aportat llum a tota la família des de la seva arribada, que m'estimo moltíssim i a qui desitjo una vida plena d'amor i d'oportunitats.

Aquesta tesi ha rebut un ajut de l'Institut de Llengües de la Universitat de Lleida per a la correcció lingüística (convocatòria de 2019)

Disseny gràfic de la portada: Àlex Moreno (www.bymore.net)

Índex

Declaració d'autoria	iii
Resum	vii
Abstract	ix
Resumen	xi
Agraïments	xiii
Índex	xv
Índex de figures	xvii
Índex de taules	xix
Llista d'abreviacions	xxi
1 Introducció	1
1.1 Tipus d'avaluacions econòmiques	2
1.1.1 Anàlisi de minimització de costos	2
1.1.2 Anàlisi de cost-efectivitat	3
1.1.3 Anàlisi de cost-utilitat	3
1.1.4 Anàlisi de cost-benefici	4
1.2 Elements clau de l'avaluació econòmica en salut	4
1.2.1 Definició de l'objectiu de l'estudi	4
1.2.2 Perspectiva d'anàlisi	5
1.2.3 Horitzó temporal d'anàlisi	5
1.2.4 Identificació de les alternatives	5
1.2.5 Estimació dels costos i dels resultats en salut	6
1.2.6 Altres aspectes rellevants	12
1.3 Mètodes d'avaluacions econòmiques	12
1.3.1 Avaluacions econòmiques basades en estudis clínics	13
1.3.2 Avaluacions econòmiques basades en models analítics	15
1.4 Anàlisi de resultats	21
1.5 Incertesa i anàlisi de sensibilitat	24
1.6 Criteris de decisió	32
1.7 Mètodes estadístics en avaluació econòmica	33

1.8	Aplicació en problemes de salut	35
1.8.1	Cribatge poblacional de càncer de mama	36
1.8.2	Prevenió cardiovascular	38
2	Hipòtesis i objectius	41
2.1	Hipòtesi general	41
2.2	Hipòtesis específiques	41
2.3	Objectius	42
3	Mètodes i resultats	43
3.1	Mètodes i resultats de l'objectiu 1	44
3.2	Mètodes i resultats de l'objectiu 2	85
3.3	Mètodes i resultats de l'objectiu 3	106
3.4	Mètodes i resultats de l'objectiu 4	120
4	Discussió	153
4.1	Resultats principals	153
4.2	Discussió conjunta	154
4.3	Limitacions	157
4.4	Fortaleses	159
4.5	Futures línies de recerca	161
5	Conclusions	163
5.1	Implicacions per a la salut pública	164
A	Altres articles amb resultats col·laterals	167
A.1	Article A1	169
A.2	Article A2	170
A.3	Article A3	171
A.4	Article A4	173
A.5	Article A5	175
A.6	Article A6	176
A.7	Article A7	177
A.8	Article A8	178
	Bibliografia	179

Índex de figures

1.1	Taxonomia de mesures de resultats en salut	9
1.2	Representació gràfica del càlcul dels anys de vida ajustats per qualitat (AVAQ)	10
1.3	Procés de selecció del model analític adequat	17
1.4	Pla de cost-efectivitat	22
1.5	Frontera òptima de Pareto	24
1.6	Diagrama de punts de l'anàlisi de sensibilitat probabilística pel programa de cribratge B respecte de l'estratègia de no cribratge	29
1.7	Corba d'acceptabilitat de cost-efectivitat per al programa de cribratge B respecte de l'estratègia de no cribratge	29
1.8	Diagrama de punts de l'anàlisi de sensibilitat probabilística de totes les alternatives de cribratge	30
1.9	Corba d'acceptabilitat de cost-efectivitat per a totes les alternatives de cribratge	30
1.10	Exemple de presa de decisions en l'exemple de les alternatives de cribratge de càncer de mama	32
1.11	Danys i beneficis del cribratge mamogràfic	37

Índex de taules

1.1	Tipus d'anàlisi en avaluació econòmica	2
1.2	Tipus de costos	7
1.3	Avantatges i inconvenients dels dissenys d'estudi d'avaluacions econòmiques	14
1.4	Resultats de l'anàlisi de cost-utilitat de programes de cribratge de càncer de mama	23
1.5	Tipus o fonts d'incertesa	26
1.6	Paràmetres i distribucions candidates recomanades	28
3.1	Relació entre els objectius de la tesi i els capítols on es detalla la metodologia i es presenten els resultats	43
A.1	Objectius col·laterals a la tesi i referències a les publicacions originals	167

Llista d'abreviacions

ASD	anàlisi de sensibilitat determinista
ASP	anàlisi de sensibilitat probabilística
AVAQ	anys de vida ajustats per qualitat
ICER	<i>incremental cost-effectiveness ratio</i>
ICP	intervenció coronària percutània
ICUR	<i>incremental cost-utility ratio</i>
MRC	malaltia renal crònica
NSTE-ACS	<i>non-ST-segment elevation acute coronary syndrome</i>
QVRS	qualitat de vida relacionada amb la salut
RCT	<i>randomized controlled trial</i>
SCA	síndrome coronària aguda
SNP	<i>single-nucleotide polymorphism</i>
STEMI	<i>ST-segment elevation myocardial infarction</i>
VEIP	valor esperat de la informació perfecta

A la meva estimada mare

Capítol 1

Introducció

La presa de decisions és el procés que resulta de l'elecció d'una acció entre diverses alternatives possibles davant d'un problema determinat o d'una situació donada. Així doncs, la presa de decisions és present en tots els aspectes de la vida de les persones, des del pla purament individual o familiar fins a disciplines socials molt diverses —negocis, finances, etc.— o polítiques públiques, com per exemple, la política sanitària.

Com que l'objectiu és realitzar una elecció òptima, o si més no prou bona, sempre caldrà identificar prèviament totes les alternatives factibles. De fet, sense identificar com a mínim una alternativa a l'escenari de referència —no fer res o seguir fent el mateix— no hi ha decisió possible a prendre. Es tracta, doncs, d'ordenar totes les alternatives possibles segons el valor (preferència, utilitat, creença) que el decisor atorga a cadascuna, tenint en compte les conseqüències que se'n deriven, tant els danys com els beneficis. També, donat que els recursos són limitats, és important avaluar quins són els costos que s'han d'assumir per tal de dur a terme l'elecció presa [1]. Tal com va definir Lionel Robins, l'«economia és la ciència que estudia el comportament humà com a relació entre els fins i els mitjans escassos que tenen usos alternatius» [2, p. 15]. Destinar els recursos d'una determinada manera implica renunciar a altres alternatives d'ús d'aquests recursos. En economia, es defineix el *cost d'oportunitat* com el valor de la millor opció a la qual es renuncia quan es realitza una elecció [3].

Ja fa temps que els raonaments econòmics s'han incorporat al camp de la salut [1, 4]. L'emergent mercat d'innovació de tecnologies sanitàries, que té com a objectiu millorar la salut de la població, també genera uns costos associats al sistema sanitari i a la societat en general. Enfront de la il·limitada demanda de salut, cal decidir quina és la millor manera d'utilitzar l'escassetat de recursos disponibles, que fins i tot pot conduir a deixar de finançar tecnologies o intervencions considerades obsoletes [5, 6]. En concret, l'avaluació econòmica té per objecte informar i facilitar decisions sobre quin és el millor ús dels recursos limitats disponibles, amb la finalitat de maximitzar-ne els beneficis [4, 7]. És a dir, ajuda a prioritzar les diferents alternatives segons

critèris d'eficiència [7]. Per això, tota avaluació econòmica mesura els costos i les conseqüències de les alternatives, i els compara per informar sobre l'elecció òptima. Així doncs, l'avaluació econòmica s'ha definit com «l'anàlisi comparativa de cursos alternatius d'acció basant-se en els seus costos i en les seves conseqüències» [7, p. 4].

1.1 Tipus d'avaluacions econòmiques

La identificació i la mesura dels costos són similars en la majoria de les avaluacions econòmiques. Per contra, la manera de mesurar les conseqüències derivades de les alternatives examinades pot variar considerablement. Aquestes diferències donen lloc a diferents tipus d'anàlisis (taula 1.1).

TAULA 1.1: Tipus d'anàlisis en avaluació econòmica

Tipus d'anàlisi	Mesura dels costos	Mesura dels resultats en salut
Anàlisi de minimització de costos	Unitats monetàries	Sense diferències en resultats en salut
Anàlisi de cost-efectivitat	Unitats monetàries	Unitats clíniques (morts evitades, reducció de pressió arterial, etc.)
Anàlisi de cost-utilitat	Unitats monetàries	Esperança i qualitat de vida (anys de vida ajustats per qualitat)
Anàlisi de cost-benefici	Unitats monetàries	Unitats monetàries (disposició a pagar)

Adaptat de Sacristán *et al.* «Evaluación económica en medicina». *Medicina Clínica (Barcelona)*, 2004, vol. 122, p. 379-382.

1.1.1 Anàlisi de minimització de costos

L'anàlisi de minimització de costos és el tipus d'anàlisi que normalment es fa servir per descriure la situació en què les conseqüències de les alternatives comparades són equivalents, independentment de les unitats en què es mesuren, tant en efectivitat com en seguretat. Així doncs, si el benefici clínic és igual, la diferència entre aquestes es redueix a una comparació de costos. L'única aplicació possible de l'anàlisi de minimització de costos és en situacions en què s'adopta una visió basada en investigacions prèvies o en l'opinió dels professionals, en què les alternatives comparades són equivalents en termes d'efectivitat, com per exemple tractaments amb bioequivalents [7]. Les veus més crítiques han reclamat l'abandó total d'aquest tipus d'anàlisi per inapropiada, atès que argumenten que no és adequat avaluar les diferències en efectes i costos de manera separada i seqüencial, alhora que defensen que les incerteses haurien de ser avaluades conjuntament per tal de no introduir biaixos en l'avaluació [8, 9].

1.1.2 Anàlisi de cost-efectivitat

L'anàlisi de cost-efectivitat és aquella en què els costos estan associats a un resultat en salut únic i comú a les alternatives comparades, i que podria resultar diferencial entre aquestes [7]. Aquest tipus d'anàlisi és molt freqüent en avaluació econòmica [10] i es caracteritza perquè els resultats en salut es mesuren en unitats utilitzades habitualment en la pràctica clínica. Els resultats en salut poden basar-se en variables intermèdies —reducció de la pressió arterial, respostes al tractament, complicacions evitades i altres— o finals —vides salvades, anys de vida guanyats— (vegeu la secció 1.2.5.2). L'ús de mesures d'efecte específiques de certa àrea clínica suposa la seva major limitació, ja que dificulta l'avaluació del cost d'oportunitat (benefici perdut potencial) en altres àrees que podrien ser cobertes amb el mateix pressupost. Aquesta limitació només pot ser superada amb l'ús de mesures genèriques de benefici en salut.

1.1.3 Anàlisi de cost-utilitat

L'anàlisi de cost-utilitat és essencialment una variant de la de cost-efectivitat, en què les conseqüències es mesuren en escales genèriques de benefici en salut. De fet, sovint hom es refereix a l'anàlisi de cost-utilitat com de cost-efectivitat. Així doncs, les conseqüències en salut es calculen ponderant el temps en diferents estats de salut per les preferències individuals o socials envers aquests, en l'intent de recollir efectes sobre diferents aspectes de la salut. En economia, el terme *utilitat*¹ és freqüentment utilitzat per referir-se a les preferències del consumidor pels diferents béns i activitats de consum. Els economistes de la salut fan servir el terme *utilitat* per referir-se al nivell subjectiu de benestar que els individus experimenten en diferents estats de salut [11]. Així, en el camp de la salut, com més preferit sigui un estat de salut determinat, major *utilitat* s'hi associa [7, 12]. Els valors de les utilitats s'expressen habitualment en una escala numèrica amb extrems en 0, que representa la mort, i 1, que representa la millor salut possible; això és l'estat de perfecte benestar físic, mental i social, i no només absència de malaltia, segons l'Organització Mundial de la Salut.² En alguns sistemes de puntuació també es poden donar valors negatius d'utilitat per a estats de salut pitjors que la mort. Els valors d'utilitat de cada un dels estats de salut poden variar d'una manera molt important d'un país a un altre, encara que s'estandarditzi per estructura de població [13]. Els valors d'utilitat permeten obtenir *anys de vida ajustats per qualitat*, que és una mesura àmpliament utilitzada en anàlisis de cost-utilitat (vegeu la secció 1.2.5.2).

¹*Utility* [Internet]. York Health Economics Consortium; 2016. <https://www.yhec.co.uk/glossary/utility/>

²Constitució de l'Organització Mundial de la Salut [Internet]. World Health Organization (WHO); 2014. <https://www.who.int/governance/eb/constitution/es/>

1.1.4 Anàlisi de cost-benefici

Teòricament, l'anàlisi de cost-benefici és la forma d'avaluació econòmica més completa i més sòlida, i ha estat utilitzada com a ajuda per a la presa de decisions en diversos àmbits de la política econòmica i social del sector públic des de mitjans del segle xx [14], però poc utilitzada per a l'avaluació econòmica de programes i serveis amb conseqüències en l'estat de la salut [15]. En l'anàlisi de cost-benefici tant els costos com els resultats en salut es mesuren en unitats monetàries. En conseqüència, és l'única que pot ser utilitzada per informar de decisions sobre assignació de recursos entre programes sanitaris i no sanitaris [16]. Tot i ser el tipus d'anàlisi preferida per molts economistes de la salut, la seva limitació principal rau en la dificultat d'obtenir estimacions, deguda a les resistències a realitzar valoracions en unitats monetàries (*disposició a pagar*) de canvis en l'estat de salut [4, 16, 17]. Entre els diferents mètodes emprats per estimar la disposició a pagar destaquen el mètode de valoració contingent i l'anàlisi conjunta (*conjoint analysis*) [15, 16, 18-20].

1.2 Elements clau de l'avaluació econòmica en salut

Valorar els resultats d'una avaluació econòmica sol ser una tasca difícil, perquè cal valorar si els mètodes utilitzats són els adequats i si els resultats són vàlids en un altre context [7]. A l'hora d'avaluar la validesa de l'evidència, normalment es revisen els mètodes emprats, per la qual cosa sovint és útil separar els diversos elements dels mètodes per tal de valorar-los cadascun de més a prop. Llavors, així com existeix la guia CONSORT, on s'indiquen els elements clau que cal reportar en assajos clínics [21]; la guia STROBE per a estudis observacionals [22], i tota una plètor de guies publicades, totes elles incloses dins el lloc web de l'EQUATOR Network (Enhancing the Quality and Transparency of Health Research) [23], també s'ha desenvolupat la guia CHEERS, amb les recomanacions per al report d'avaluacions econòmiques [24]. A continuació s'identifiquen els elements clau de qualsevol avaluació econòmica que un lector o revisor pot trobar-se en avaluacions ben realitzades, i que els permetran identificar i avaluar els punts forts i febles de qualsevol estudi.

1.2.1 Definició de l'objectiu de l'estudi

Com en qualsevol estudi, sempre cal formular clarament l'objectiu, que en avaluació econòmica ha d'especificar clarament quines són les alternatives comparades, la perspectiva a partir de la qual es fa la comparació, l'horitzó temporal en què s'examinen tant els costos com les conseqüències de les alternatives comparades, i les característiques de la població d'estudi [7].

1.2.2 Perspectiva d'anàlisi

Un dels aspectes més rellevants és el de la perspectiva de l'avaluació econòmica. La perspectiva més global és la de la societat, encara que existeixen altres possibilitats com són la perspectiva del sistema públic de salut, la d'una institució en particular —per exemple, un hospital o un centre de salut—, la del professional sanitari responsable de la intervenció, o la del pacient i la seva família. L'elecció d'una perspectiva en concret determina els costos rellevants, així com la manera en què aquests s'han d'avaluar. Així doncs, mentre que la perspectiva del sistema sanitari restringeix els costos a aquells que corresponen als serveis de salut prestats pel sistema en qüestió, l'adopció de la perspectiva de la societat implica incloure tota inversió de recursos que pugui suposar un cost d'oportunitat en qualsevol aspecte de la societat [12]. En darrera instància, la perspectiva de l'estudi depèn de qui el financi; així, mentre que hi ha països on es recomana com a preferent la perspectiva del finançador —usualment, el sistema nacional de salut—, opcionalment acompanyada de manera complementària per la perspectiva social [25-29], aquesta és la recomanada com a preferent en altres regions [30-33]. Sempre que es considerin diferents perspectives, els resultats s'han de presentar de manera separada i diferenciada per a cadascuna.

1.2.3 Horitzó temporal d'anàlisi

L'elecció de l'horitzó temporal és una consideració metodològica important en les avaluacions econòmiques. Hauria de ser prou llarg per recollir les principals conseqüències econòmiques i en salut —tant els beneficis com els danys— diferencials entre les alternatives comparades [34]. Un horitzó temporal curt pot incorporar biaixos en la comparació de les alternatives, si en aquestes els costos i els resultats en salut es produeixen a un ritme diferent [12]. Per tant, per a la majoria de les avaluacions econòmiques, l'horitzó temporal pertinent és tota la vida del pacient [7].

Si bé els models de decisió poden ampliar-se considerant futures cohorts d'incidents i l'abast de la valoració es pot estendre molt de temps en el futur, això inevitablement comporta estimacions cada cop més incertes. Per tant, sovint és convenient limitar l'anàlisi a les cohorts més immediates per a les quals es necessiten decisions polítiques. Així mateix, l'omissió de resultats futurs llunyans pot arribar a ser negligible a causa de la taxa de descompte, una correcció sobre els resultats que s'incorpora en l'avaluació econòmica per reflectir les preferències temporals i de la qual es tracta amb més profunditat a la secció 1.2.5.3. La recomanació general és que els analistes reflexionin detingudament sobre com diferents horitzons temporals poden incidir en els resultats i que justifiquin la seva elecció [35].

1.2.4 Identificació de les alternatives

Informar d'una decisió concreta requereix identificar els possibles cursos alternatius que es podrien prendre per millorar la salut dels pacients que es

troben en una situació determinada. Tot i que pot suposar un enorme treball i pot complicar l'avaluació, en principi, s'han de tenir en compte totes les alternatives rellevants —totes aquelles que tinguin alguna possibilitat de ser eficients, que pot incloure *no fer res*—, ja que una avaluació excessivament limitada, en la qual no s'hagi considerat alguna alternativa rellevant, pot ser enganyosa. Les alternatives poden incloure diferents combinacions o seqüències de tractament i diferents maneres en què es pot utilitzar una intervenció. Generalment, les alternatives solen comparar-se amb la pràctica clínica habitual o amb l'opció més efectiva [29, 33, 36].

Sovint, les decisions sobre quina alternativa oferir es poden prendre per a diferents tipus de pacients que, tot i tenir la mateixa condició, probablement difereixen en els efectes de les intervencions. En aquestes situacions, cal identificar els diferents subgrups de pacients i les alternatives dins de cada subgrup. Les estimacions dels costos, dels danys i dels beneficis sobre la salut per a tota la població de pacients són la suma dels costos i dels efectes en salut en cadascun dels subgrups [7].

1.2.5 Estimació dels costos i dels resultats en salut

L'estimació dels costos i dels resultats en salut és un aspecte clau que sosté la validesa de qualsevol avaluació econòmica en salut. Com s'assenyalava en un editorial del *British Medical Journal*, presentar l'evidència de l'efectivitat clínica d'una determinada intervenció sense cap evidència sobre els costos que suposa seva aplicació «és com un aparador de botiga sense preus» [37]. Igualment, les avaluacions econòmiques en salut no han de donar-se sense una base d'evidència científica idealment publicada sobre els resultats en salut de les alternatives comparades. Així, tots dos conceptes són igualment importants.

1.2.5.1 Identificació, quantificació i valoració dels costos

L'anàlisi de costos i la comparació entre les diferents alternatives són comunes a tots els tipus d'avaluació econòmica. El cost d'una intervenció és el producte de dos elements: la quantitat total de recursos utilitzats i el valor monetari de les unitats dels recursos en qüestió. Per determinar el cost d'una intervenció és habitual seguir tres fases successives: identificació, quantificació i valoració [7, 12].

Tradicionalment, els costos han estat classificats segons si estan relacionats directament amb la intervenció o no (directes o indirectes, respectivament) i segons si impliquen un servei sanitari o no. Per tal d'identificar quins costos s'han d'incloure en l'avaluació econòmica cal establir-ne la perspectiva, així com l'horitzó temporal (taula 1.2).

Un cop identificats, cal determinar la quantitat de recursos consumits en unitats físiques. És possible valorar els recursos en diferents nivells d'agregació. Per exemple, es pot estimar el cost d'una hospitalització com els dies d'estada

TAULA 1.2: Tipus de costos

	Sanitaris	No sanitaris
Directes	Tractaments farmacològics, serveis hospitalaris, serveis ambulatoris, etc.	Ajudes tècniques, modificacions a la llar, transport, cuidadors formals, etc.
Indirectes	Serveis sanitaris no relacionats amb la intervenció	Pèrdua/reducció de productivitat (absentisme/presentisme), cuidadors informals, etc.

Adaptat de Prieto *et al.* «Análisis de costes y resultados en la evaluación económica de las intervenciones sanitarias». *Medicina Clínica (Barcelona)*, 2004, vol. 122, p. 423-429.

a l'hospital pel cost d'un dia d'ingrés, que inclouria tots els recursos hospitalaris (temps del personal assistencial, tractaments farmacològics, restauració, allotjament, etc.), o es pot identificar i quantificar tots aquests recursos. El nivell d'agregació dels recursos ha de ser consistent amb l'avaluació en qüestió, o pot veure-se'n limitada la validesa. Naturalment, el nivell d'agregació sempre estarà supeditat a la informació disponible [12]. Un aspecte important al qual cal prestar atenció i que cal evitar és el *dobte recompte* d'un mateix recurs.

Finalment, cal valorar els recursos assignant un *cost unitari* equivalent al seu cost d'oportunitat. Teòricament són els preus de mercat els que valoren adequadament el cost d'oportunitat, així que generalment són aquests els emprats. Alguns països, com el Regne Unit i Austràlia, disposen d'una llista oficial de tarifes nacionals [38, 39]. La disponibilitat i l'accés a un repositori únic i oficial de costos unitaris permeten la generalització i la comparació de resultats entre diferents estudis; simplifiquen i agilitzen l'accés a la informació i, per tant, del cost de realitzar l'avaluació econòmica, i minimitzen el risc de biaixos en les estimacions [3, 32]. Tanmateix, els repositoris únics també plantegen inconvenients, com pot ser la manca de representativitat en avaluacions específiques per un centre o, en el cas d'Espanya, per comunitat autònoma [32].

Entre tots els tipus de costos, la inclusió dels costos sanitaris indirectes és el que actualment planteja major controvèrsia. No està exempta de certa controvèrsia la inclusió dels costos indirectes no sanitaris en una anàlisi amb una perspectiva més àmplia de societat, ja sigui en la mateixa definició o bé en com han de ser avaluats. Aquest fet es reflecteix en l'alta variabilitat en les recomanacions de les diferents agències de tecnologies sanitàries. Si bé totes les agències inclouen els costos directes sanitaris, i moltes també consideren els costos directes no sanitaris, són menys les que també inclouen els costos indirectes en la perspectiva social. Entre aquestes, n'hi ha que només consideren la pèrdua de productivitat, mentre que d'altres també inclouen el

cost del temps dels familiars [40]. Els costos indirectes corresponents a cuidadors informals són cada vegada més rellevants en els països desenvolupats, donat l'elevat nombre de pacients amb malalties cròniques, característica demogràfica pròpia d'aquests països. En conseqüència, les atencions informals en termes de costos i morbiditat dels cuidadors creixen en pes relatiu sobre el conjunt dels costos [41]. En la majoria d'avaluacions que inclouen els costos indirectes es fa servir el *mètode del capital humà*, tot i que també es fa servir l'aproximació coneguda com a *costos de fricció* [7].

El mètode del capital humà estima els costos de la pèrdua de productivitat mitjançant el salari brut percebut per l'individu desocupat a causa de la malaltia. S'acostuma a fer servir taules salarials mitjanes, com les publicades per les oficines nacionals d'estadística, i no el salari de l'individu. Una raó és la incomoditat que acostuma a provocar preguntar pels ingressos i, per tant, la poca fiabilitat de les respostes. Una altra raó, no menys important, és per consideracions d'*equitat*, ja que, assumint salaris específics, certs programes socials destinats a població amb salaris baixos podrien resultar ineficients davant d'alternatives de les quals es podria beneficiar una població més benestant. Alguns estudis també atorguen un cost per a aquells que no tenen feina remunerada —per exemple, les persones encarregades de les tasques de la seva pròpia llar—, ja sigui aplicant salaris mitjans, el cost de contractació per substituir la tasca de l'individu o el cost d'oportunitat de la producció a què haurien pogut contribuir si no s'estiguessin a casa [7].

El mètode dels costos de fricció estima la quantitat de producció perduda a causa de la malaltia sobre la base del temps que les organitzacions necessiten per restaurar el nivell inicial de producció. Aquest període de fricció difereix segons ubicació, indústria, empresa i categoria de treballadors. Per exemple, tal com diuen Drummond *et al.* [7] «pot caldre només mitja jornada per formar un cambrer d'una cadena d'hamburgueseries, però almenys dos dies per formar un economista de salut!». Les estimacions obtingudes amb aquest mètode són substancialment menors que les obtingudes amb el mètode del capital humà [7, 42]. L'alta variabilitat en els resultats dels costos indirectes segons el mètode emprat en dificulta la inclusió en les avaluacions econòmiques [32].

Finalment, també es poden identificar els costos intangibles, tot i que no se'n recomana la valoració en termes monetaris degut a l'enorme complexitat que suposa quantificar econòmicament aspectes com l'ansietat, la por, etc., que poden patir els individus. Tanmateix, en anàlisis de cost-utilitat, en les quals els resultats en salut es mesuren en escales de qualitat de vida relacionada amb la salut (QVRS), aquestes poden incloure l'impacte en aspectes intangibles dins la qualitat de vida autopercibuda.

1.2.5.2 Mesures de resultats en salut

L'elecció de la mesura de resultat en salut a emprar ve determinada per la pregunta d'investigació i pel tipus d'avaluació que es preveu realitzar (vegeu

la secció 1.1). En la literatura s'hi pot trobar un gran nombre de mesures diferents de l'efecte en salut de les intervencions (figura 1.1).

Sovint, les mesures obtingudes en estudis clínics no són realment mesures del resultat en salut en si, sinó de resultats clínics intermedis —per exemple, un control lipídic o de pressió arterial millor en el risc cardiovascular. Tot i que aquestes mesures de resultats intermedis poden ser adequades per determinar si una intervenció és més efectiva que una altra, per si mateixes no poden indicar la magnitud de la millora en salut que proporciona una intervenció. Per fer-ho, els canvis en el resultat intermedi han d'estar vinculats a canvis en les mesures del resultat en salut en si mateix.

Potser el principal problema de la majoria de les mesures clíniques de resultats és que són específiques de l'àrea clínica en qüestió —per exemple, reducció d'atacs d'epilèpsia, dies lliures de símptomes de l'asma. Poden ser útils en avaluacions econòmiques restrictives, en què cal escollir entre teràpies per a la mateixa condició, però continua plantejant dificultats d'interpretació. Per exemple, si una teràpia és més costosa que una altra, però proporciona més millora clínica, no és clar si aquesta millora compensa el cost addicional. Per poder fer aquesta interpretació cal suposar que la mesura utilitzada inclou

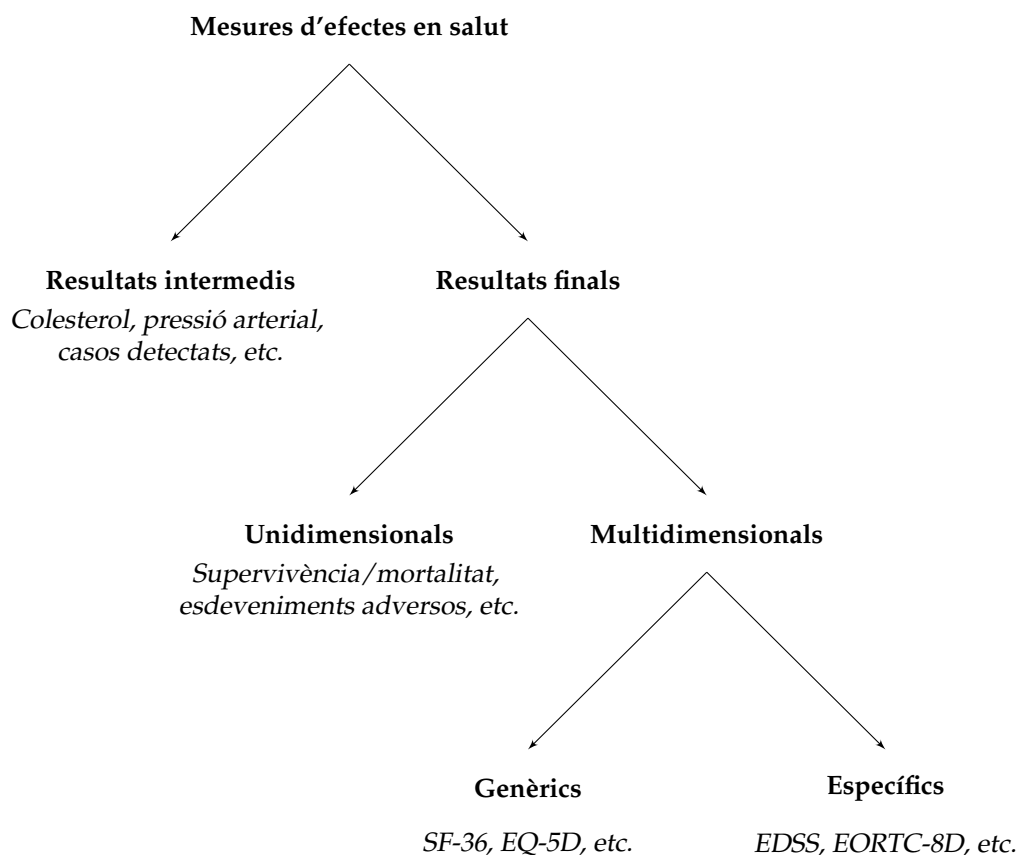


FIGURA 1.1: Taxonomia de mesures de resultats en salut

Adaptat de Drummond *et al.* *Methods for the economic evaluation of health care programmes*. Oxford University Press, 2015, p. 28.

tots els canvis de salut rellevants [7, 43].

Algunes mesures del resultat en salut es limiten a un sol aspecte de la salut, com per exemple, els efectes sobre la mortalitat i la supervivència. Tot i que la durada de la vida és clarament un aspecte important de la salut, la qualitat en què es viu també és important. De fet, en el cas de teràpies contra el càncer amb alta toxicitat, el pacient pot enfrontar-se a un compromís entre la durada de la vida i la qualitat de vida. Per tant, una avaluació econòmica basada en el cost per any de vida guanyat és, en el millor dels casos, incompleta i també podria ser enganyosa [7]. Atès que les mesures de qualitat de vida se centren en els efectes del tractament que, en principi, repercuteixen en el benestar del pacient, val la pena considerar-ne la rellevància per a l'avaluació econòmica. Aquest fet ha motivat un interès creixent a incloure mesures de resultats informats pel pacient en estudis clínics, que principalment consisteixen en mesures de satisfacció i de QVRS dels pacients i que recullen aspectes de l'efecte del tractament que potser no es recopilen amb els principals resultats clínics.

Moltes mesures de QVRS són específiques de determinades malalties i intenten recollir aquells aspectes de la salut que són més importants per als pacients amb una condició particular —per exemple, l'*Expanded Disability Status Scale* en esclerosi múltiple. En aquest sentit, pateixen limitacions similars als resultats clínics. Les mesures genèriques consideren un ampli ventall de dimensions de qualitat de vida que, en principi, podrien ser afectades per qualsevol malaltia, incloent-hi la funció física, el benestar mental, la funció social i el dolor —per exemple, l'*SF-12* o l'*EQ-5D*. Així, fins i tot la tasca aparentment senzilla de descriure la salut planteja qüestions difícils sobre quines

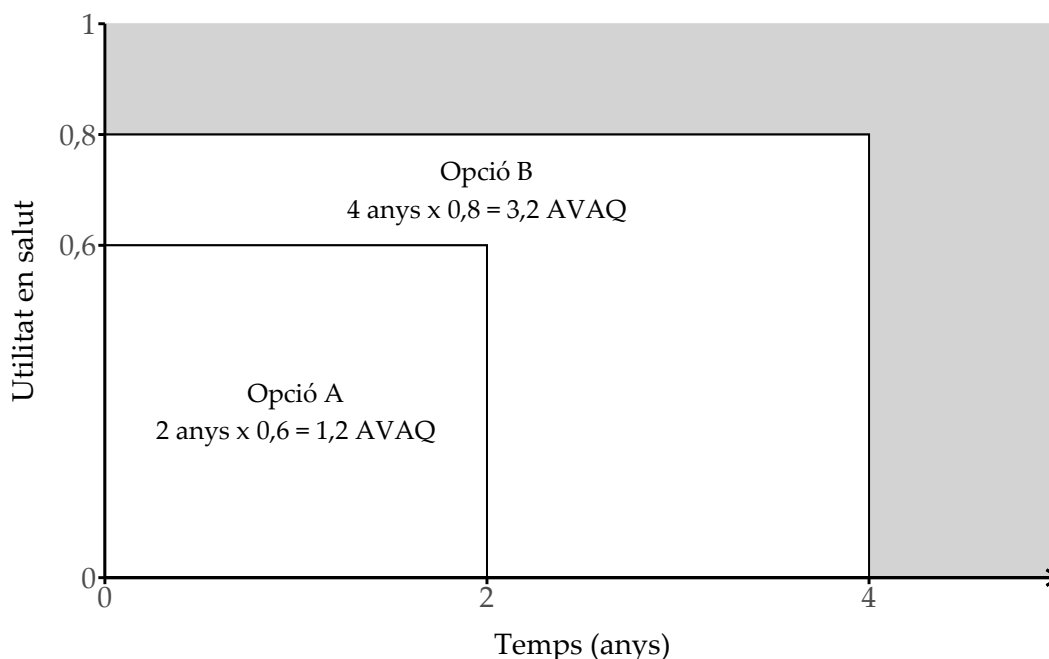


FIGURA 1.2: Representació gràfica del càlcul dels anys de vida ajustats per qualitat (AVAQ)

dimensions o atributs són importants. Hi ha diversos motius pels quals una mesura genèrica de la QVRS té avantatges, fins i tot quan hi ha un sol resultat clínic rellevant de la teràpia, com l'augment de la supervivència. Per una banda, permet comparar els efectes en salut que proporcionen alternatives que afecten diferents aspectes de la salut o tenen efectes en diverses malalties. Per altra banda, proporciona coherència amb com es prenen altres decisions rellevants per a altres grups de pacients amb diferents malalties.

La mesura genèrica més àmpliament utilitzada en anàlisis de cost-utilitat és els *anys de vida ajustats per qualitat* (AVAQ). Els AVAQ es calculen simplement multiplicant els anys viscuts amb un cert estat de salut pel pes de la QVRS (és a dir, la utilitat en salut; vegeu la secció 1.1.3) associat a aquest estat de salut (figura 1.2). Malgrat la seva aparent simplicitat, els AVAQ permeten combinar la quantitat i la qualitat de vida de les persones en un únic índex de resultat que permet les comparacions entre diferents malalties o àrees de salut [7]. S'han proposat alternatives als AVAQ, com els *anys saludables equivalents* [44], els *anys de vida ajustats per discapacitat* [45] i, més recentment, mesures que tracten l'estat de salut amb atributs més amplis de capacitat de benestar [46], o donant més pes al component de morbiditat [47].

Una altra crítica als AVAQ és que reflecteixen una perspectiva individual i no una perspectiva social, ja que tots es valoren igual, tot i que en l'àmbit social es podrien valorar diferent segons la situació [7]. Per exemple, es podria voler valorar més un benefici en AVAQ d'algú que pateix una malaltia greu, en contraposició a una de més lleu, o valorar-ne més alt el benefici en la població més jove. S'han proposat les *vides joves salvades equivalents*, com a alternativa als AVAQ convencionals [48], com a mesura de les preferències socials. Mentre que en l'enfocament dels AVAQ, les preferències s'obtenen preguntant pels intercanvis o per les compensacions (*trade-offs*) que voldrien per a ells mateixos, en l'enfocament de les vides joves salvades equivalents, que es basa en mètodes d'intercanvis personals (*person trade-off*), les preferències s'obtenen preguntant a cada membre de la societat quins tipus d'intervencions voldrien per a altres, i això constitueix la base per a la presa de decisions de la societat [7]. El debat sobre si els AVAQ reflecteixen les preferències de la societat planteja qüestions sobre si han de ser ponderats per considerar altres factors —com l'equitat— en les decisions sobre l'assignació de recursos assistencials [7]. En general, es prefereix no introduir ponderacions en els AVAQ en la mateixa anàlisi («un AVAQ és un AVAQ» [49, 50]) i deixar la consideració de pertinència d'aquests factors en el procés deliberatiu de la presa de decisions [7].

1.2.5.3 Descompte de costos i resultats en salut futurs

De la mateixa manera que els resultats en salut poden ocórrer en diferents períodes temporals, és habitual que les alternatives comparades suposin assumir costos en diferents moments del temps. Els exemples més clars d'aquest fet són aquells en els quals es compara una intervenció preventiva —vacunes, detecció precoç, etc.— enfront d'intervencions terapèutiques. Mentre que les intervencions terapèutiques proporcionaran resultats en salut

a curt termini, les intervencions preventives suposaran despeses immediates però els resultats en salut s'observaran a llarg termini. Normalment es prefereixen els beneficis a curt termini i que els danys o els costos es posterguin en el temps. Els economistes anomenen aquest fenomen «preferència temporal» [1, 7, 12] i succeeix, en part, per la incertesa del futur, i també per raons econòmiques, ja que en períodes de creixement econòmic positiu (inflacionistes) un euro d'avui té més valor que un euro en el futur.

La preferència temporal ha d'incorporar-se a l'avaluació econòmica introduint ajustos sobre els resultats econòmics i en salut futurs que permetin expressar-los en el seu valor actual, és a dir, el seu valor equivalent en el moment present. Aquesta operació d'ajust consisteix a multiplicar les quantitats futures per un factor de descompte, que depèn de la taxa de preferència temporal i del temps que transcorre fins al moment en què es contreu el cost o es gaudeix del benefici. Hi ha una certa controvèrsia sobre quina ha de ser la taxa de descompte, i sobre si cal aplicar la mateixa taxa als costos i als resultats en salut [1, 12]. La taxa de descompte a aplicar sovint és la indicada en les directrius oficials de les agències nacionals d'avaluació de tecnologies sanitàries de cada país, entre les quals existeix una considerable variabilitat [40]. Les directrius actuals de diversos països es poden consultar a l'eina en línia desenvolupada per la International Society for Pharmacoeconomics and Outcomes Research [51]. La recomanació estàndard és fer servir taxes que facilitin la comparació entre estudis, típicament del 3% o del 3,5% anual, tant en costos com en beneficis [7].

1.2.6 Altres aspectes rellevants

Existeixen altres elements que són clau a l'hora de dissenyar l'avaluació econòmica i que els que la reben o llegeixen han de saber identificar i valorar. Així, l'origen de les dades i el mètode utilitzat, l'obtenció dels resultats incrementals i l'anàlisi de la incertesa són també aspectes sobre els quals cal fer una revisió crítica per tal d'avaluar si els mètodes emprats en l'estudi són adequats i els resultats, vàlids. Si els resultats són vàlids, també cal valorar com s'aplicarien per a la presa de decisions [7]. En les seccions següents es tracta d'aquests aspectes: es relaten breument les alternatives metodològiques emprades en avalució econòmica segons l'origen de les dades sobre les quals es basa, la presentació de resultats i l'anàlisi de la incertesa associada, i de quina manera es poden fer servir per informar d'una decisió.

1.3 Mètodes d'avaluacions econòmiques

En l'avaluació econòmica, una de les dificultats rau en l'accés a la informació, per a la identificació tant de les alternatives com de les conseqüències que derivarien de cadascuna en cas de ser finalment escollida. La manca d'informació rellevant explícita només pot ser coberta per acords tàcits o assumpcions. Tanmateix, com que la salut és probablement el bé més preuat

de les persones, la presa de decisions en salut ha d'estar basada en la major i millor evidència disponible, d'estudis ben dissenyats i ben realitzats. Així, les dades necessàries per realitzar una avaluació econòmica es poden obtenir d'assaigs clínics aleatoritzats (RCT, de l'anglès *randomized controlled trial*) i d'estudis observacionals. Les diferències entre els dos tipus de disseny d'estudi tenen implicacions pels tipus de dades que se'n poden obtenir i per la interpretació de les conclusions resultants [52]. Principalment, els RCT són la font preferent d'eficàcia (resultats en condicions ideals), mentre que els dissenys observacionals aporten resultats en condicions reals i, per tant, proporcionen informació d'efectivitat.

És poc probable que un únic estudi proporcionï tota la informació necessària per estimar els costos i els efectes en salut de totes les alternatives rellevants, durant el període de temps en què probablement es diferenciïn, o en diferents subgrups de pacients. Per tant, el més habitual és haver de cercar sistemàticament l'evidència disponible. Els models analítics s'utilitzen comunament com l'estructura en què l'evidència de diferents fonts, en combinació amb supòsits, pot informar dels paràmetres que proporcionen estimacions de costos i efectes sobre la salut en qualsevol període de temps per a totes les alternatives [7]. La taula 1.3 resumeix els avantatges i inconvenients dels possibles dissenys d'estudi per a una avaluació econòmica.

1.3.1 Avaluacions econòmiques basades en estudis clínics

La principal fortalesa dels RCT prové de l'alta validesa interna. No obstant això, els RCT tenen limitacions substancials [53]. La generalització dels resultats es veu habitualment limitada degut a criteris restrictius de selecció de pacients —per exemple, pacients en estat sever, dones no embarassades, etc.— que els poden diferenciar en característiques de pronòstic i en els efectes de les intervencions de la població de referència potencialment tributària [7, 54]. A banda, els RCT determinen els efectes de les intervencions seguint protocols en condicions experimentals ideals; per tant, reporten mesures d'eficàcia. Ara bé, una avaluació econòmica té com a objectiu relacionar costos i efectes en condicions reals; per tant, hauria d'incorporar mesures d'efectivitat [7, 52, 54]. Altres limitacions dels RCT per fer avaluacions econòmiques tenen a veure amb el temps de seguiment i amb els comparadors. Mentre que idealment una avaluació econòmica inclou resultats al llarg de la vida en comparació amb la pràctica clínica habitual, els RCT acostumen a tenir temps de seguiment més aviat curts i sovint comparant la nova intervenció o tractament amb placebo [7, 54]. Tot i aquestes limitacions, els RCT han estat la font d'informació necessària per justificar decisions d'aprovació, d'entrada al mercat i de finançament de noves tecnologies en el sistema sanitari. Així, la International Society for Pharmacoeconomics and Outcomes Research publica una guia amb recomanacions per a realitzar avaluacions econòmiques en assajos clínics [55, 56]. Algunes de les seves limitacions poden ser superades, com a mínim en part, amb el disseny d'RCT pragmàtics, que miren d'avaluar els efectes d'una intervenció en condicions més semblants a les de la rutina clínica habitual i amb un seguiment prolongat dels pacients [7, 54, 57].

TAULA 1.3: Avantatges i inconvenients dels dissenys d'estudi d'avaluacions econòmiques

Tipus d'estudi	Avantatges	Inconvenients
Assaig clínic	Alta validesa interna Alt nivell d'evidència (fiabilitat dels resultats)	Resultats d'eficàcia (no d'efectivitat) Població molt restringida (manca de representativitat poblacional) Seguiments curts Ús de recursos condicionats per protocol Cost elevat
Estudi observacional	Resultats d'efectivitat Major representativitat poblacional Costos i resultats a llarg termini Menor cost que els assajos clínics	Baixa validesa interna Menor nivell d'evidència que un assaig clínic Qualitat de les dades
Model analític	Incorporació de dades d'efectivitat Horitzons temporals llargs Comparació de diverses alternatives Baix cost i ràpida execució	Incorporació d'assumpcions Complexitat Crítiques de manca de transparència

Adaptat de Rubio-Terrés *et al.* «Métodos utilizados para realizar evaluaciones económicas de intervenciones sanitarias». *Medicina Clínica (Barcelona)*, 2004, vol. 122, p. 578-583.

En els estudis observacionals els pacients reben els tractaments i segueixen les intervencions partint de la pràctica clínica habitual, sense o amb poques restriccions de protocol, la qual cosa els ha valgut el nom d'estudis del «món real» [52]. La seva principal limitació rau en la manca d'aleatorització del tractament, fet que pot limitar la comparació de grups o inclús impossibilitar-la [7]. Per exemple, si certa intervenció només s'ofereix a un subgrup de pacients en estat greu, mentre que la resta de pacients són sotmesos a una altra intervenció, no hi ha cap base sobre la qual poder comparar l'efectivitat d'ambdues intervencions. Per ajustar el desequilibri en les característiques relacionades amb la selecció del tractament i potencialment predictives dels resultats, cal fer servir modelització estadística, que proporciona la metodologia necessària amb què es pot arribar a estimar l'efecte causal de les intervencions en dissenys observacionals [58, 59]. Tanmateix, els dissenys observacionals superen diverses de les limitacions dels RCT per a l'avaluació econòmica. Així, els estudis observacionals solen tenir temps de seguiment més llargs de cohorts més grans i més representatives de la població.

Els dissenys prospectius sempre seran preferibles als retrospectius, en què la manca d'informació sobre les característiques o esdeveniments clínics pot limitar potencialment l'avaluació econòmica. En els darrers anys, hi ha hagut un gran interès en l'ús per a la investigació d'informació recollida de manera rutinària dins els sistemes de salut, també d'aquella amb finalitats administratives, incloent-hi pagaments per part dels proveïdors de salut. Alguns sistemes de salut estan intentant enllaçar una sèrie de conjunts de dades administratives i clíniques, en ocasions integrades amb noves dades recopilades per altres proveïdors, per proporcionar conjunts de dades que cobreixen un gran nombre de pacients en un ampli ventall d'àrees clíniques [7]. En aquest sentit, el sistema sanitari públic a Catalunya desenvolupa el Pla Director de Sistemes d'Informació, que dona resposta a les necessitats d'informació de les persones i dels professionals que actuen en tot el sistema de salut i en col·laboració amb els àmbits que contribueixen a la salut i al benestar de les persones, com el social o l'educatiu [60].

1.3.2 Avaluacions econòmiques basades en models analítics

Un enfocament més general per a l'avaluació econòmica es dona quan s'extreuen evidències de diferents fonts dins d'un model analític de decisions. L'ús dels models matemàtics en problemes de salut pública té una llarga història, des de les contribucions durant el segle XVIII del pioner Daniel Bernoulli en la modelització de la propagació i la mortalitat per verola [61], fins a treballs posteriors apareguts a cavall dels segles XIX i XX, principalment relacionats amb malalties infeccioses [62-65]. En els darrers anys, el nombre d'estudis de modelització matemàtica ha augmentat notablement, i tracten d'un ampli ventall de preguntes. Cada vegada més, l'enfocament de la medicina basada en l'evidència i, per tant, la presa de decisions clíniques, es basa en la caracterització matemàtica o estadística de processos biològics [66]. No obstant això, els estudis de modelització rarament s'inclouen formalment com a part de la jerarquia de l'evidència. L'Organització Mundial de la Salut actualment està desenvolupant criteris d'avaluació del grau de credibilitat dels estudis de modelització per tal d'integrar l'evidència que se'n deriva amb dades primàries per informar de directrius i recomanacions en polítiques de salut [67].

Sovint en avaluació econòmica s'ha de recórrer a l'ús de models matemàtics per superar les limitacions existents en els resultats dels estudis clínics. A més a més, en ocasions no és factible endegar un assaig clínic o un estudi de cohorts prospectives, sovint per raons de caire ètic o pressupostari. Els models matemàtics, ben dissenyats i ben realitzats tot i que amb les seves pròpies limitacions, són una bona eina per aportar evidència. Un model analític de decisió en salut defineix un conjunt de relacions matemàtiques entre estats de salut i/o processos terapèutics, que caracteritzen el ventall de possibles pronòstics de malalties i l'impacte de les intervencions alternatives en costos i efectes sobre la salut que interessin per a l'avaluació econòmica [7]. Així, l'ús de models analítics es fa necessari per: extrapolar els resultats de cost i d'efectivitat més enllà de les dades observades en un assaig clínic;

vincular resultats clínics intermedis amb resultats finals; generalitzar els resultats obtinguts en un assaig a la pràctica clínica habitual; generalitzar els resultats aplicant particularitats dels sistemes de salut de diferents regions o països; realitzar comparacions entre alternatives per a les quals no hi ha comparacions directes en estudis clínics; quan cal estimar la càrrega global de certa malaltia o condició [67], o quan cal informar de les decisions sobre polítiques d'assignació de recursos que s'han de prendre en absència de *dades dures* [68], com pot ser en el cas d'avaluar l'eficiència de la vacuna, encara per desenvolupar, per a una mutació d'un virus o l'aparició d'un nou virus, com el SARS-CoV-2, que està causant una epidèmia o una pandèmia.

Els models permeten fer simulacions de realitats complexes i incertes tot extrapolant, adaptant i integrant els resultats dels assajos clínics utilitzant informació procedent de registres hospitalaris, bases de dades, opinions d'experts, revisió d'històries clíniques i estudis epidemiològics, entre d'altres [69]. La qualitat i la credibilitat dels models dependran directament de la validesa de la informació emprada en la seva realització i de la seva transparència i reproductibilitat [54, 68]. S'han dut a terme diverses iniciatives per a la definició de guies de bones pràctiques per al desenvolupament i l'avaluació dels models econòmics de salut. Per exemple, la International Society for Pharmacoeconomics and Outcomes Research i la Society for Medical Decision Making publiquen conjuntament una guia amb recomanacions per a realitzar avaluacions econòmiques basades en models analítics d'intervencions sanitàries [70]. També, les agències d'avaluació de tecnologies sanitàries publiquen les seves respectives recomanacions, amb una considerable variabilitat entre si, però sobre la modelització només argumenten vagament que el model adequat depèn del problema de decisió [40].

En la literatura s'hi troben diverses propostes de marc conceptual per seleccionar el tipus de model apropiat [71-75]. La figura 1.3 és una adaptació de l'esquema proposat per Barton *et al.* [71], amb la incorporació d'altres tipus de models dels quals no tracta en el seu treball. El procés de selecció d'un enfocament adequat per als models econòmics sanitaris implica la consideració de diversos criteris. Un aspecte comú en tots els esquemes proposats és considerar aspectes estructurals com el nivell d'agregació de la població —simulació de cohorts enfront de microsimulació— i la interactivitat de la població. A més a més d'alinear la naturalesa d'un problema de decisió determinat amb les característiques estructurals d'un tipus de modelització, cal considerar possibles restriccions pràctiques que inclouen des de la disponibilitat de dades fins a la comprensió de les necessitats dels usuaris finals i les parts interessades. Una recomanació general és seleccionar el model més senzill que descriu adequadament els estats de salut sense fer supòsits poc raonables, atès que els models senzills i transparents són més fàcilment entesos i acceptats per no especialistes [73]. Tanmateix, l'ús de marcs de decisió pot proporcionar una falsa sensació de confiança en el fet que s'utilitzen els mètodes adequats sense aprofundir en el procés i les consideracions que han dut a l'elecció d'un cert enfocament de modelització sobre els altres [76]. L'elecció del tipus de modelització és un pas important i necessari, amb grans

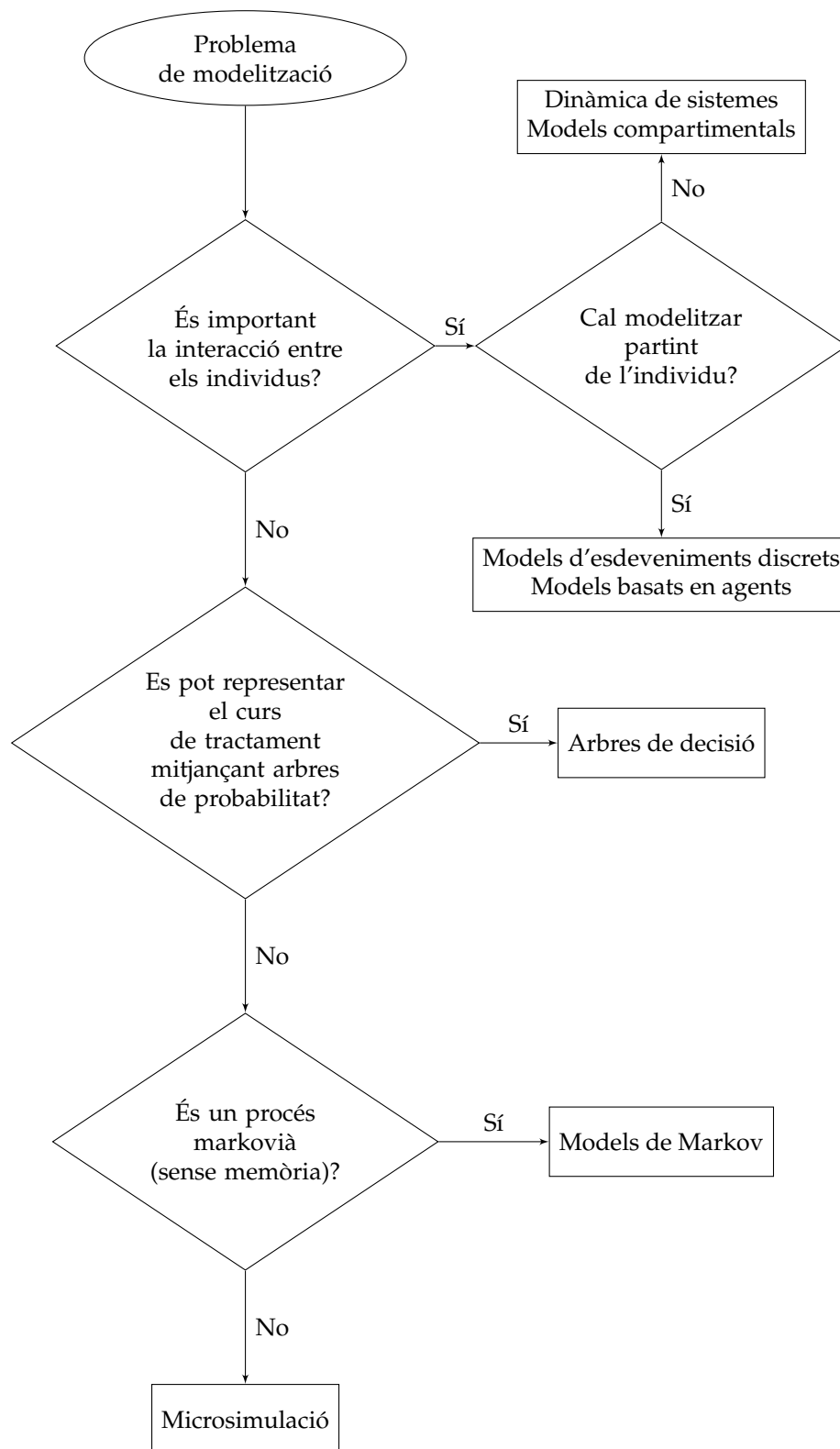


FIGURA 1.3: Procés de selecció del model analític adequat

Adaptat de Barton *et al.* «Modelling in the economic evaluation of health care: selecting the appropriate approach». *Journal of Health Services Research & Policy*, 2004, vol. 9, p. 110-118.

implicacions en el desenvolupament i en la validesa del model posterior, raons per les quals cal triar rigorosament, i debatre i informar de l'elecció acuradament [75-77].

Els arbres de decisió són probablement el tipus de modelització amb l'estructura més simple i familiar, en què tots els possibles cursos del pacient es mostren explícitament amb probabilitats associades i mesures de resultats de costos i de salut. Els costos i els resultats en salut esperats de les alternatives comparades es poden calcular sumant al llarg de tots els cursos possibles dels pacients el cost i el resultat en salut de cada curs ponderats per la seva respectiva probabilitat [7]. Quan s'utilitzen arbres, s'assumeix que la població que s'està examinant es pot modelitzar agregadament, de manera que les probabilitats coincideixen amb la proporció d'individus de la població que s'està estudiant [74]. En general, solen ser adequats si el període de temps és curt i si la mortalitat dels pacients no difereix entre les estratègies [71]. No permeten considerar recurrències, però són possibles escenaris molt complicats, amb moltes estratègies de decisió, seqüències de tractament llargues i múltiples resultats possibles [72]. Així, els arbres es poden estendre per més d'un període de temps, però després d'unes poques repeticions l'arbre pot tenir molts centenars de nodes [73]. Per tant, els models de simulació i de Markov són més convenients per modelitzar intervencions amb efectes en salut a llarg termini.

Les limitacions dels arbres de decisions són la raó principal per la qual els models de Markov també s'utilitzen àmpliament en l'avaluació econòmica en salut [7]. Mentre que els arbres de decisió caracteritzen possibles pronòstics en diferents branques de l'arbre, els models de Markov es basen en una sèrie d'estats de salut que un pacient pot ocupar en un moment determinat. Els estats de salut són mútuament excloents, representen la totalitat del procés de la malaltia i s'assumeix que els pacients que es troben en un mateix estat de salut formen una població homogènia. La probabilitat que un pacient ocupi un determinat estat es valora en una sèrie de períodes de temps discrets, anomenats cicles. La durada d'aquests cicles depèn de la malaltia i de les intervencions que s'estiguin avaluant: per exemple, pot ser d'un mes o d'un any, però idealment ha de ser tan curt com sigui possible en sentit clínic. El moviment entre estats es produeix un cop per cicle segons les probabilitats de transició. Les probabilitats de transició depenen només de l'estat en què es troba el pacient a l'inici del cicle (assumpció de Markov de no-memòria).

L'assumpció de Markov no permet que la probabilitat de transició depengui de la història prèvia, sinó que només depèn de l'estat actual. D'aquesta manera, els models de Markov obliguen a crear estats separats per a cada subgrup, amb atributs o característiques diferents [72, 73]. El model s'executa en molts cicles, i se n'obté la distribució de pacients en els estats de salut al llarg del temps, fins que tota la cohort està en l'estat absorbent de la mort o fins que s'assoleix l'horitzó temporal del model. Cada estat del model generalment té un cost associat, així com una mesura de resultat en salut, tot i que també es poden aplicar a la proporció que realitza una transició. Per a cada cicle es pot calcular la suma ponderada dels resultats de costos i de salut

de cada estat per la proporció de la cohort en els respectius estats de salut. Els costos i els resultats en salut totals s'obtenen sumant-los al llarg dels cicles de l'horitzó temporal del model. En la seva forma més general, un model de Markov permet que les probabilitats variïn amb el temps. En aquest cas, els models s'anomenen semi-Markov i habitualment cal solucionar-los mitjançant mètodes de simulació numèrica [74]. El principal benefici dels models de Markov és la fàcil representació d'esdeveniments recurrents, però no permeten la interacció entre individus [71, 78].

Tant els arbres de decisió com els models de Markov descrits són exemples de models de cohorts [7], tot i que la metodologia es pot estendre i se'n pot fer ús per simular pacients individuals sense interacció; els més habituals són els models de microsimulació o simulació de primer ordre de Montecarlo [7, 72, 78]. Com que els individus es modelitzen per separat, la microsimulació pot emmagatzemar informació sobre el que li ha passat a l'individu, és a dir, té memòria, i tampoc cal assumir homogeneïtat entre els pacients. Les probabilitats de transició modelitzen el pronòstic del pacient i es calculen condicionades pels factors de risc i per la història individual. El càlcul de costos i resultats en salut és similar als models de Markov de cohorts, amb la particularitat que amb els models de microsimulació cada pacient té les seves pròpies estimacions de cost i resultats segons el seu recorregut pel model.

Una alternativa als models de microsimulació són els models d'esdeveniments discrets [7]. En els models d'esdeveniments discrets se simula el temps fins al proper esdeveniment per a cada pacient; per tant, un pacient pot romandre en un estat determinat durant una durada variable del temps, que no ha de ser un múltiple exacte d'un cicle de longitud fixa, com en els models de Markov. Les històries individuals se simulen una a una o totes alhora. La simulació simultània en els models d'esdeveniments discrets permet modelitzar interaccions o competència pels recursos [76]. Tots els altres tipus de models anteriors suposen que els individus són independents els uns dels altres respecte de la seva salut, és a dir, la salut d'un individu no afecta la salut d'altres. Aquesta hipòtesi d'independència pot ser insostenible en el context de malalties infeccioses en què la incidència de noves infeccions depèn del nombre d'individus infectats [7, 71], o inclús quan les restriccions en els recursos comporten que l'elecció del tractament d'un pacient afecta el que es pot donar a un altre [71].

La simulació d'esdeveniments discrets és un tipus de modelització molt flexible que descriu el flux d'entitats dins un sistema [73, 74, 76, 79]. Les entitats són objectes, com ara individus, que poden interaccionar amb altres entitats o competir pels recursos del sistema. En aquest tipus de models els recursos es representen mitjançant objectes que proporcionen serveis a les entitats dinàmiques, que són els individus, la qual cosa implementa els efectes de les cues. Cada interacció entre entitats o amb els recursos és un esdeveniment i fa variar l'estat del sistema per complet [74]. A les entitats se'ls pot donar atributs, com ara característiques o memòria, que poden influir en el seu recorregut a través de la simulació o la durada entre els esdeveniments. Les conseqüències en costos i en efectes en salut es poden associar a l'ús de

recursos, als esdeveniments o al temps amb una condició particular. Els models basats en agents podrien considerar-se una extensió dels models d'esdeveniments discrets en els quals els agents tenen coneixement del seu estat i segueixen regles de decisió per interactuar o comunicar-se amb altres agents o amb el mateix sistema. De simples regles individuals, en poden emergir comportaments complexos dins la xarxa d'agents [74, 76]. Tot i la flexibilitat dels models de simulació d'esdeveniments discrets, que els permet adaptar-se d'una manera molt més precisa a la realitat que es pretén explicar, el seu ús no és gaire habitual, probablement degut a la contrapartida en l'augment dels paràmetres necessaris per al model. Una altra raó que pot explicar el seu poc ús és que l'augment de la complexitat, encara que sigui adequada al problema, pot semblar que en redueix la transparència i dificulta als revisors comprendre el model i verificar que és una representació raonable i correctament implementada [80].

No sempre és necessari modelitzar cada individu per separat per considerar interaccions. La dinàmica de sistemes i els models compartimentals, un enfocament basat en cohorts comú en epidemiologia i investigació operativa, modelitza l'estat del sistema en termes de variables contínues canviant al llarg del temps [72, 81]. Aquests tipus de models permeten que els canvis en el sistema siguin en funció de l'estat del mateix sistema (*feedback* o retroalimentació). Aquesta retroalimentació és típica en contextos de malalties infeccioses, en què elevats nivells d'infecció comporten un major risc d'incidència de noves infeccions, tot i que també redueix el nombre de persones en risc, alhora que pot tenir efectes en el funcionament del servei de salut, que té un rendiment diferent quan és ple o se sobrepassen les capacitats d'atenció. Un model de dinàmica de sistemes s'acostuma a començar representant el diagrama causal de tots els factors rellevants i les vies potencials de retroalimentació. Aquests diagrames consten de dos tipus de variables principals: l'estoc, que representa les entitats de capacitat limitada o cues, i el flux, que són les taxes d'intensitat en les quals els estocs es buiden o es reomplen [74]. Es poden estimar costos i resultats, que s'actualitzen contínuament segons el temps en estoc o els moviments entre estocs.

En definitiva, un bon modelitzador triarà el mètode de simulació que compleixi millor la finalitat de l'estudi, tenint en compte la disponibilitat de dades i els compromisos entre la càrrega computacional, la capacitat de realitzar anàlisis de sensibilitat, la capacitat de comprendre el comportament del model i la capacitat de comunicar els resultats a l'audiència a qui va adreçada l'avaluació. Així, donat que els models d'avaluació econòmica són útils per proporcionar als decisors evidència sobre les conseqüències de les alternatives considerades, la credibilitat del model i la confiança en els resultats són ineludiblement fonamentals. Eddy *et al.* [82] descriuen les pràctiques òptimes a les quals s'haurien d'adherir els autors d'un model per tal d'aportar la informació necessària que permeti avaluar la credibilitat del model i els resultats. Principalment distingeixen dos aspectes als quals cal adreçar-se per aconseguir-ho: la transparència i la validació. La transparència del model

té el propòsit de proporcionar una descripció general del model que permeti entendre'n el funcionament, així com proporcionar la informació tècnica necessària per tal de poder reproduir-lo. Amb la validació el que es pretén és avaluar si el model prediu la realitat que intenta imitar amb exactitud. L'anàlisi de sensibilitat és una part important de la validació, però no la substitueix, ja que tot i que es pot utilitzar per explorar com canvien els resultats d'un model en funció de la variació dels paràmetres (vegeu la secció 1.5), per si mateixa no avalua amb quina precisió un model simula la realitat. Amb tot, la transparència i la validació estan inextricablement vinculades i ambdues són necessàries per ajudar els lectors a guanyar confiança en els resultats del model [82].

1.4 Anàlisi de resultats

L'anàlisi i la comparació entre les alternatives dels resultats obtinguts depenen del tipus d'avaluació econòmica. Així, en l'anàlisi de minimització de costos l'alternativa més eficient serà aquella amb un cost menor ja que, tal com s'ha explicat en la secció 1.1.1, en aquest tipus d'avaluació econòmica s'assumeix que les alternatives no difereixen en efectivitat. En l'anàlisi de cost-benefici els resultats poden presentar-se com una relació entre els beneficis i els costos, és a dir, benefici dividit per cost, ambdós expressats en les mateixes unitats monetàries, tot i que és més habitual i es recomana ordenar les alternatives segons les diferències entre els beneficis i els costos, això és benefici menys cost. Un resultat de benefici net positiu implica que l'aplicació de la intervenció incrementaria el benestar global, raó per la qual hauria de ser adoptada sempre que les restriccions pressupostàries ho permetessin. En cas d'avaluar diverses alternatives mútuament excloents, la regla de decisió és escollir en primer lloc aquella que mostri un major benefici net [1].

En l'anàlisi de cost-efectivitat es pot expressar el valor relatiu d'una intervenció com el quocient que s'obté de dividir el cost per l'efectivitat, que rep el nom de *relació de cost-efectivitat mitjana* [36]. Un altre indicador possible en una anàlisi de cost-efectivitat és la *relació de cost-efectivitat incremental* (ICER, de l'anglès *incremental cost-effectiveness ratio*), que compara els costos i els resultats d'efectivitat de dues alternatives A i B de la manera següent:

$$\text{ICER} = \frac{\text{Cost}_B - \text{Cost}_A}{\text{Efectivitat}_B - \text{Efectivitat}_A} = \frac{\Delta\text{Cost}}{\Delta\text{Efectivitat}}$$

Per tant, l'ICER informa del cost addicional per unitat d'efectivitat addicional que suposa passar de A a B. En l'anàlisi de cost-utilitat, els resultats de les opcions comparades es presenten de la mateixa manera que en l'anàlisi de cost-efectivitat, amb la particularitat que la mesura d'efectivitat són AVAQ, i els indicadors solen anomenar-se *relació de cost-utilitat mitjana* i *relació de cost-utilitat incremental* (ICUR, de l'anglès *incremental cost-utility ratio*).

Tant en l'anàlisi de cost-efectivitat com en l'anàlisi de cost-utilitat, totes les guies advoquen per la presentació de l'ICER o l'ICUR [40]. El motiu rau en

el fet que l'avaluació econòmica pretén proporcionar informació d'eficiència comparada entre les diferents alternatives, és a dir, estimar els recursos addicionals que s'han d'invertir en una intervenció a canvi d'obtenir un millor resultat en salut en comparació amb la resta d'alternatives. Per tant, la informació verdaderament rellevant l'aporten les relacions incrementals, ja que tenen en compte les diferències en costos i resultats.

Les alternatives comparades se solen representar en el *pla de cost-efectivitat* (figura 1.4). Habitualment, l'alternativa de referència, per exemple la pràctica clínica habitual, se situa al centre de coordenades. D'aquesta manera, l'eix d'abscisses representa la diferència en efectivitat, mentre que l'eix d'ordenades representa la diferència en costos. Els eixos identifiquen quatre quadrants, que es poden identificar com nord-est (NE), nord-oest (NO), sud-oest (SO) i sud-est (SE). En cas d'alternatives en els quadrants NO o SE, l'elecció és clara. En el quadrant SE l'alternativa proporciona millores en salut a menor cost, i s'anomena alternativa *dominant*. En el quadrant NO succeeix just el contrari, i les alternatives en aquest quadrant s'anomenen *dominades*. En els quadrants NE i SO, l'elecció depèn del llindar en l'ICER segons, respectivament, si s'està disposat a invertir més recursos a canvi de millors resultats en salut o si s'assumeix una pèrdua de salut a canvi d'un estalvi de recursos, que podrien ser destinats a altres fins. Arribats a aquest punt, determinar quina quantitat addicional de recursos s'ha d'invertir en una determinada alternativa per tal d'assolir una millora en salut és competència únicament de l'agent regulador i dels decisors en salut [33].

Per tal d'il·lustrar el càlcul i la interpretació de l'ICER/ICUR, considerem els resultats d'una anàlisi de cost-utilitat de programes de cribatge poblacional de càncer de mama alternatius (taula 1.4).

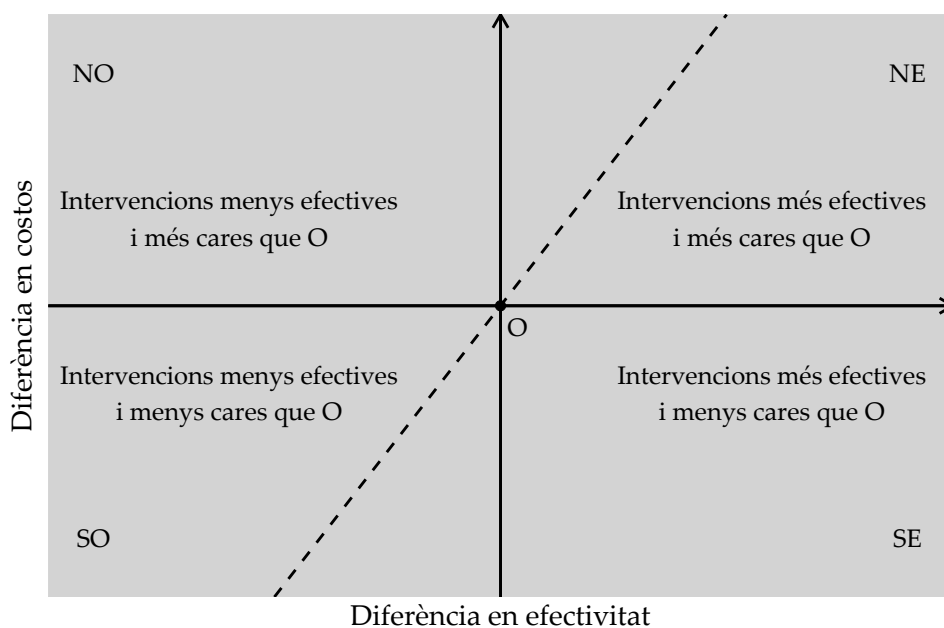


FIGURA 1.4: Pla de cost-efectivitat

TAULA 1.4: Resultats de l'anàlisi de cost-utilitat de programes de cribratge de càncer de mama

	Cost (€)	Efectivitat (AVAQ)	Δ Cost	Δ Efectivitat	ICUR
No cribratge	11.000	20	-	-	-
Programa A	12.500	20,2	1.500	0,2	7.500
Programa B	13.500	20,1	1.000	-0,1	Dominada
Programa C	14.500	20,35	2.000	0,15	13.333,33
Programa D	17.500	20,4	3.000	0,05	Dominada estesa
Programa E	20.500	20,5	6.000	0,15	40.000

Cost i AVAQ per dona. Δ : increment; AVAQ: anys de vida ajustats per qualitat; ICUR: incremental cost-utility ratio.

Les alternatives considerades són mútuament excloents, i també es considera l'opció de no cribratge, que és l'alternativa que comporta menys cost. Els diferents programes alternatius proposen diferents rangs d'edat i periodicitats de proves de cribratge que proporcionen millores en efectivitat, mesurada en AVAQ, a canvi d'un augment en la inversió de recursos.

En la taula 1.4 també es mostra el cost i l'efectivitat incremental respecte de l'alternativa que la precedeix en eficiència. Per això cal establir si hi ha alternatives dominades o si totes són dominants [1]. Així, respecte de l'alternativa de no cribratge, el programa B proporciona un guany menor en efectivitat amb un cost major que el programa A; per tant, el programa B és una alternativa dominada pel programa A. Com que el programa B queda exclòs de les alternatives eficients, l'ICUR del programa C es calcula respecte del programa A, que és l'alternativa més eficient que el precedeix en costos incrementals. També pot succeir que certa alternativa estigui dominada per una combinació d'unes altres dues, la qual cosa s'anomena *dominància estesa*. En l'exemple, el programa D no pertany al conjunt d'alternatives eficients perquè, respecte del programa C, la seva ICUR (60.000 €) és major que la del programa E (40.000 €). Equivalentment, una alternativa està en dominància estesa si l'efectivitat que proporciona es podria assolir adoptant una combinació de dues alternatives eficients a un cost menor. En l'exemple, l'efectivitat que proporciona el programa D es pot assolir oferint el programa C a 2/3 de la població, i el programa E al 1/3 restant de la població, ja que $2/3 \times 20,35 + 1/3 \times 20,5 = 20,4$ AVAQ; alhora que es reduirien els costos del programa D, ja que $2/3 \times 14.500 \text{ €} + 1/3 \times 20.500 \text{ €} = 16.500 \text{ €} < 17.500 \text{ €}$. L'exclusió del programa D de les alternatives òptimes també és un bon exemple per il·lustrar la conveniència de l'ICER (ICUR) enfront de les relacions de cost-efectivitat (cost-utilitat) mitjançant per seleccionar les alternatives eficients. La relació de cost-utilitat mitjana del programa D ($\approx 857,8 \text{ €}$) és menor que la del programa E (1.000 €), però, com s'acaba de veure, el programa D queda exclòs de les alternatives eficients per estar en situació de dominància estesa.

Per tant, les alternatives que s'haurien de tenir en compte per al seu millor rendiment d'eficiència formen la *frontera d'eficiència òptima de Pareto*, i en l'exemple són l'alternativa de no cribratge i els programes A, C i E. Totes les

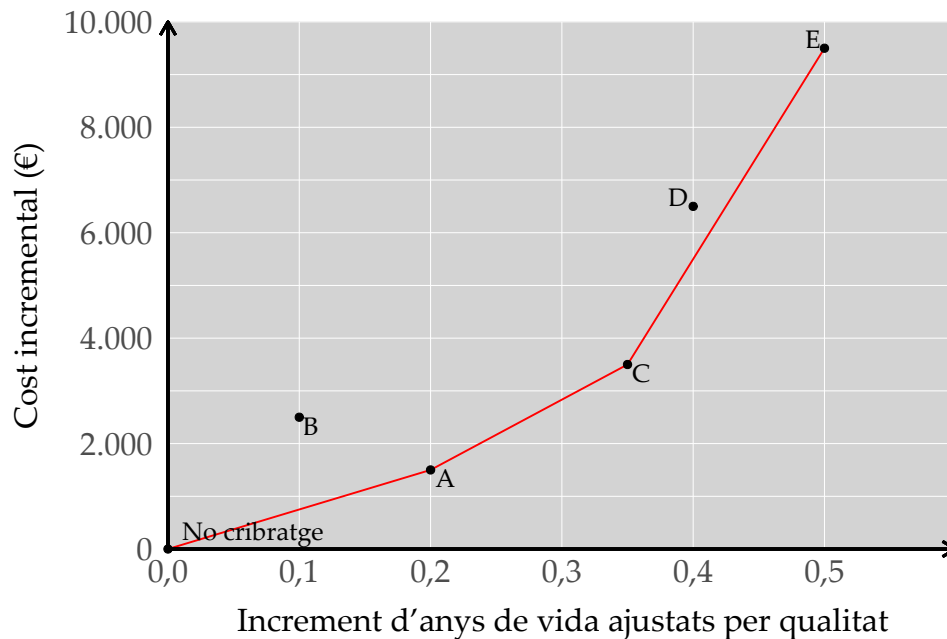


FIGURA 1.5: Frontera òptima de Pareto

alternatives de la taula 1.4 es representen en la figura 1.5, on també es poden observar les situacions de dominància descrites i la frontera d'eficiència òptima.

Existeixen alternatives a l'ICER per decidir si una intervenció és o no rendible, com el *benefici net en salut incremental* i el *benefici monetari net incremental*, definits de la manera següent:

$$\text{Benefici net en salut incremental} = \Delta\text{Efectivitat} - 1/\lambda \times \Delta\text{Cost}$$

$$\text{Benefici monetari net incremental} = \lambda \times \Delta\text{Efectivitat} - \Delta\text{Cost}$$

on λ és el *llindar d'eficiència*, o el màxim que s'està disposat a pagar per benefici en salut incremental. Si una alternativa és rendible, tant el benefici net en salut incremental com el benefici monetari net incremental són positius, que és equivalent al fet que l'ICER sigui inferior al llindar λ que estableixi l'agent decisor. En un sistema de salut amb recursos restringits, els costos assistencials representen realment els resultats en salut d'altres pacients en competència sobre els recursos assistencials. Per tant, les decisions basades en l'avaluació econòmica tracten d'identificar l'alternativa que ofereix els majors beneficis nets incrementals [7].

1.5 Incertesa i anàlisi de sensibilitat

Tota avaluació econòmica conté un cert grau d'incertesa, imprecisió i contrària metodològica. Per això, cal identificar els supòsits metodològics crítics i les fonts d'incertesa, i quantificar-ne l'impacte en els resultats per tal de

complementar la informació per a la presa de decisions, tal com indiquen acadèmics, societats científiques i agències d'avaluació estatals [3, 40, 55, 83, 84]. En el cas d'avaluacions basades en models analítics de decisió, en què es fa servir l'evidència de diverses fonts per tal de determinar els paràmetres del model, resulta imprescindible mesurar la incertesa deguda als mètodes analítics utilitzats i la variabilitat de l'evidència disponible, així com avaluar l'extrapolació i la generalització dels resultats a altres poblacions o a la població general [85]. Briggs *et al.* (2012) [86] distingeixen entre diversos tipus o fonts d'incertesa, relacionats amb les dades utilitzades en l'anàlisi o els principals supòsits metodològics. En concret, la incertesa estocàstica es diferencia de la incertesa paramètrica, de l'heterogeneïtat i de la incertesa estructural associada als mètodes analítics o estadístics utilitzats (taula 1.5).

Els mètodes per tractar i quantificar la incertesa difereixen segons la seva font i el tipus d'avaluació econòmica que es faci. Així, la incertesa estocàstica es relaciona amb la variabilitat no explicada, és a dir, amb el fet que diferents individus amb les mateixes característiques i probabilitats experimentin diferents efectes d'una malaltia o intervenció. En canvi, la incertesa paramètrica es refereix a la variabilitat en l'estimació del paràmetre d'interès, ja que les probabilitats que governen els resultats són en si mateixes incertes. La distinció entre la incertesa estocàstica i la paramètrica és anàloga a la diferència entre la desviació estàndard i l'error estàndard en les estimacions. Com per a l'error estàndard, la incertesa paramètrica pot ser reduïda augmentant el nombre de mesures; i anàlogament a la desviació estàndard, la incertesa estocàstica no pot ser reduïda sinó únicament pot ser caracteritzada, per exemple, descrivint la funció de densitat o la probabilitat acumulada. Per exemple, en el cas d'avaluacions econòmiques realitzades simultàniament amb un assaig clínic, les dades són estocàstiques, és a dir, tenen una distribució. Així, en les avaluacions econòmiques basades en l'anàlisi de dades individuals, els mètodes estadístics permeten avaluar tant la incertesa estocàstica com la incertesa paramètrica. Per al càlcul dels intervals de confiança de les ICER s'han proposat diverses aproximacions —mètodes basats en caixes de confiança o el·lipses de confiança, sèries de Taylor, mètode de Fieller, remostreig per *bootstrap*—, dels quals el mètode paramètric basat en el teorema de Fieller i l'enfocament no paramètric de *bootstrap* són els que proporcionen millors resultats [85, 87].

L'heterogeneïtat fa referència a la variabilitat deguda a diferències en les característiques dels pacients, com poden ser l'edat o el sexe, per exemple. Per tractar l'heterogeneïtat cal identificar els subgrups de pacients, per als quals s'haurien de realitzar anàlisis d'eficiència diferenciades. D'aquesta manera, les alternatives òptimes poden diferir entre diferents subgrups; alhora, la ponderació dels resultats per la distribució dels subgrups permet obtenir resultats poblacionals [7, 86].

Tot i que l'estructura general d'un model analític de decisió resulta de molts supòsits i opcions analítiques (vegeu la secció 1.3.2), és útil distingir dues grans categories de models que reflecteixen l'estructura subjacent i com es relacionen amb els conceptes d'incertesa [86]. En els models de simulació

TAULA 1.5: Tipus o fonts d'incertesa

Tipus d'incertesa	Descripció
Incertesa estocàstica (error de Montecarlo de primer ordre)	Variabilitat aleatòria entre pacients idèntics
Incertesa paramètrica (error de Montecarlo de segon ordre)	Variabilitat en l'estimació del paràmetre d'interès (error estàndard)
Heterogeneïtat	Variabilitat entre subgrups de pacients
Incertesa estructural (incertesa del model)	Assumpcions del model del decisió

Adaptat de Briggs *et al.* «Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6». *Value in Health*, 2012, vol. 15, p. 835-842.

pel que fa al pacient —per exemple, models d'esdeveniments discrets i de microsimulació—, la incertesa de primer ordre està relacionada amb l'error de simulació (simulació de primer ordre de Montecarlo; vegeu la secció 1.3.2) i es pot eliminar simulant un gran nombre d'històries individuals fins que l'efecte de la incertesa estocàstica sobre els resultats esdevingui insignificant. Avaluat la incertesa paramètrica, o de segon ordre, requereix l'eliminació de la incertesa estocàstica. En altres paraules, en aquests models cal simular tant la incertesa en les estimacions dels paràmetres com la variabilitat estocàstica, fet que resulta computacionalment molt intens, i, en conseqüència, sovint no s'avalua de manera minuciosa. S'han proposat alternatives metodològiques computacionalment més eficients, tot i que encara són metodologies en desenvolupament i d'ús poc estès [88-90]. Per contra, en els models de cohorts, com per exemple els models de Markov, es pot avaluar la incertesa dels paràmetres sense preocupació per la incertesa estocàstica, raó per la qual són preferibles davant dels models de microsimulació, sobretot en el cas que el decisor no estigui interessat a conèixer els resultats d'un individu en concret [88]. En avaluacions basades en models analítics, el mètode per quantificar la incertesa paramètrica s'anomena «anàlisi de sensibilitat», en què es varien els diversos paràmetres del model per tal d'informar de com aquestes variacions afecten els resultats de l'estudi [7]. L'anàlisi de sensibilitat també s'utilitza per avaluar la incertesa relacionada amb supòsits metodològics, que alhora és un complement clau per a la validació dels models analítics de decisió [82].

A l'hora de realitzar una anàlisi de sensibilitat cal identificar els paràmetres amb incertesa, especificar els rangs plausibles d'aquests paràmetres i escollir el tipus d'anàlisi de sensibilitat adient [7]. En principi, tots els paràmetres haurien de ser inclosos en l'anàlisi de sensibilitat. Gairebé l'única raó per excloure'n un paràmetre és que se'n conegui el valor amb absoluta seguretat, com per exemple podria ser el cost unitari d'un recurs en una certa jurisdicció

en concret. També cal justificar els intervals i les distribucions plausibles assignades als paràmetres, ja que assumint variacions injustificadament petites es pot concloure erròniament un alt grau de robustesa dels resultats obtinguts [7]. La incertesa pot ser avaluada mitjançant una *anàlisi de sensibilitat determinista* (ASD) o mitjançant una *anàlisi de sensibilitat probabilística* (ASP) [86]. En una ASD s'avalua la influència de la incertesa en un o més paràmetres sobre els resultats esperats, variant els paràmetres al llarg d'un rang de valors plausibles, per exemple, l'interval de confiança del 95% basat en els resultats d'un estudi. En una ASP, preferiblement tots els paràmetres es varien simultàniament a partir de distribucions de probabilitats definides a priori.

Històricament, la forma més habitual d'anàlisi de sensibilitat ha estat l'ASD univariable, i se n'han representat els resultats en diagrames de tornado. Avui en dia no es considera un mètode que permeti tractar de manera integral la incertesa dels paràmetres, ja que no incorpora la variabilitat combinada en diversos paràmetres ni la possible correlació existent. Una variant de l'ASD univariable és l'*anàlisi de llindars*, amb què es pretén identificar els valors crítics, o punts de tall, per sobre o per sota dels quals les conclusions de l'avaluació canvien a favor d'una o altra alternativa comparada. Per exemple, els responsables de prendre una decisió poden especificar una ICER per sobre de la qual el programa no seria acceptable. Llavors, es poden valorar quines combinacions de paràmetres podrien fer que se superés el llindar. Alternativament, es poden definir els valors llindar dels paràmetres que farien que el programa no fos rendible. Llavors, els responsables de la decisió després podrien decidir si els llindars es podrien o no incomplir [7]. Un plantejament més elaborat és l'ASD multivariable, que preveu incertesa en més d'un paràmetre alhora, cadascun dels quals variant dins un interval especificat. Per pocs paràmetres que hi hagi, l'ASD multivariable i la presentació de resultats es compliquen inevitablement degut al gran nombre de combinacions potencialment possibles. Una alternativa és plantejar diversos valors discrets dels paràmetres que descriguin diferents escenaris, entre els quals típicament s'inclouen el millor i el pitjor dels escenaris possibles [7, 85, 86]. L'*anàlisi d'escenaris* també es pot fer servir per explorar l'impacte d'assumpcions del model analític.

Univariable o multivariable, l'ASD no pot reflectir la incertesa combinada de tots els paràmetres del model, alhora que proporciona estimacions esbiaixades en models no lineals [7, 89]. L'ASP ben realitzada és un enfocament més rigorós, que supera les limitacions de l'ASD, fet que l'ha convertit en el tipus d'anàlisi de sensibilitat actualment més estès, alhora que recomanat en la majoria de guies i per les agències d'avaluació [7, 25, 27, 40, 86]. Si l'estimació dels paràmetres està basada en un gran volum de dades disponibles, pel teorema del límit central, la distribució normal pot ser emprada en l'ASP, de la mateixa manera que per establir regions de confiança en l'ASD. No obstant això, s'acostumen a especificar distribucions segons els tipus de paràmetre.

La taula 1.6 mostra les distribucions recomanades per als paràmetres més comuns. Ja sigui per manca d'evidència publicada o de dades disponibles, de

TAULA 1.6: Paràmetres i distribucions candidates recomanades

Paràmetre (suport)	Tipus de dades i mètodes d'estimació	Distribució candidata
Probabilitat ($0 < \theta < 1$)	Binomial: proporció Multinomial: proporcions Temps fins a l'esdeveniment: anàlisi de supervivència	Beta Dirichlet Log-normal
Risc relatiu ($\theta > 0$)	Binomial: ràtio de proporcions	Log-normal
Cost ($\theta \geq 0$)	Suma ponderada recursos: mitjana	Gamma Log-normal
Desutilitat ($\theta \geq 0$)	Contínua: mitjana	Gamma Log-normal
Tots els paràmetres (amb dades suficients)	Qualsevol distribució	Normal

Adaptat de Briggs *et al.* *Decision modelling for health economic evaluation*. Oxford University Press, 2006, p. 108.

vegades hi ha molt poca informació sobre un paràmetre i cal recórrer a l'opinió d'experts. En cap cas s'han d'excloure els paràmetres d'una anàlisi de sensibilitat per manca d'informació suficient per estimar la incertesa. S'han proposat mètodes per obtenir distribucions de probabilitat partint de l'opinió d'experts [91]. Distribucions possibles però poc plausibles, com per exemple la uniforme o la triangular, haurien de ser evitades [86].

La representació dels resultats d'una ASP depèn de si hi ha múltiples alternatives. Quan només es comparen dues alternatives, simplement es pot representar els costos i els resultats esperats simulats a l'ASP en el pla de cost-efectivitat. Reprenent l'exemple de les alternatives de cribratge de càncer de mama de la secció 1.4, la figura 1.6 mostra l'ASP de considerar el programa B i l'estratègia de no cribratge com a alternativa de referència. L'ICUR esperada és un quocient dels costos i AVAQ incrementals esperats, no una mitjana de les ICUR simulades. L'ICUR del programa B respecte del no cribratge és de 25.000€ per AVAQ. El llindar de rendibilitat està representat per la línia discontinua, amb un pendent de 20.000€ per AVAQ. Per tant, el programa B no és rendible i el no cribratge proporcionaria beneficis nets esperats superiors. La incertesa sobre la decisió es pot aproximar per la proporció de punts que es troben per sota de la línia discontinua, en què el programa B és rendible i proporciona beneficis nets superiors que el no cribratge. Ara bé, la incertesa de la decisió és en si mateixa incerta. Així, variant el llindar de rendibilitat es pot construir una corba, anomenada «corba d'acceptabilitat de cost-efectivitat», en què a l'eix d'ordenades es representa la probabilitat de rendibilitat per a un rang d'ICER a l'eix d'abscisses (figura 1.7).

Habitualment hi ha més de dues alternatives, i el diagrama de punts sobre el pla de cost-efectivitat esdevé impossible d'interpretar correctament. Les

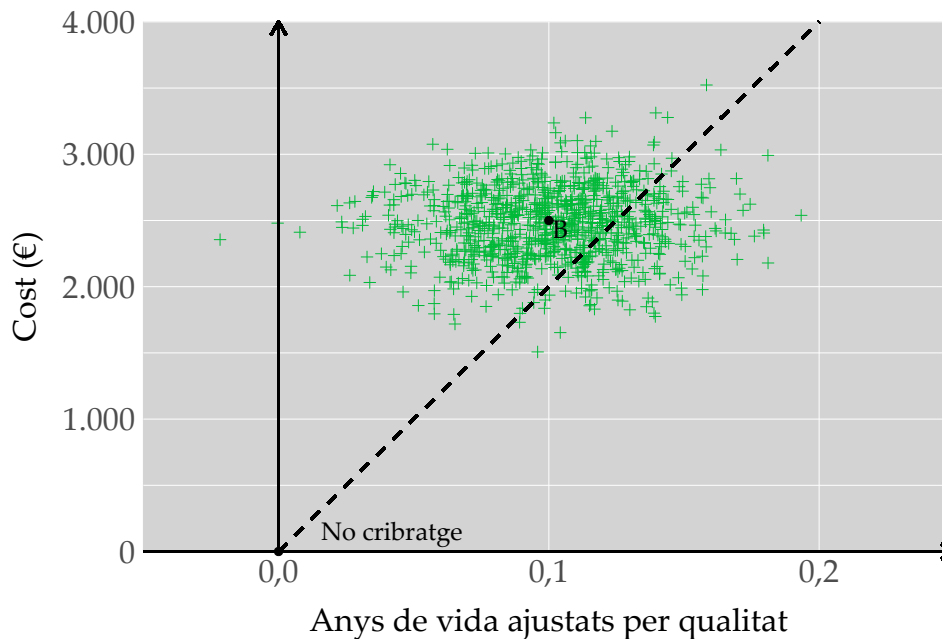


FIGURA 1.6: Diagrama de punts de l'anàlisi de sensibilitat probabilística pel programa de cribratge B respecte de l'estratègia de no cribratge

corbes d'acceptabilitat permeten avaluar més fàcilment els resultats de l'anàlisi de sensibilitat. En les figures 1.8 i 1.9 es mostren els resultats de l'anàlisi de sensibilitat de totes les alternatives de l'exemple del cribratge de càncer de mama. En el cas de múltiples alternatives, les corbes d'acceptabilitat s'obtenen a partir del benefici net esperat calculat per a diferents valors del llindar de rendibilitat. La probabilitat que cadascuna de les alternatives

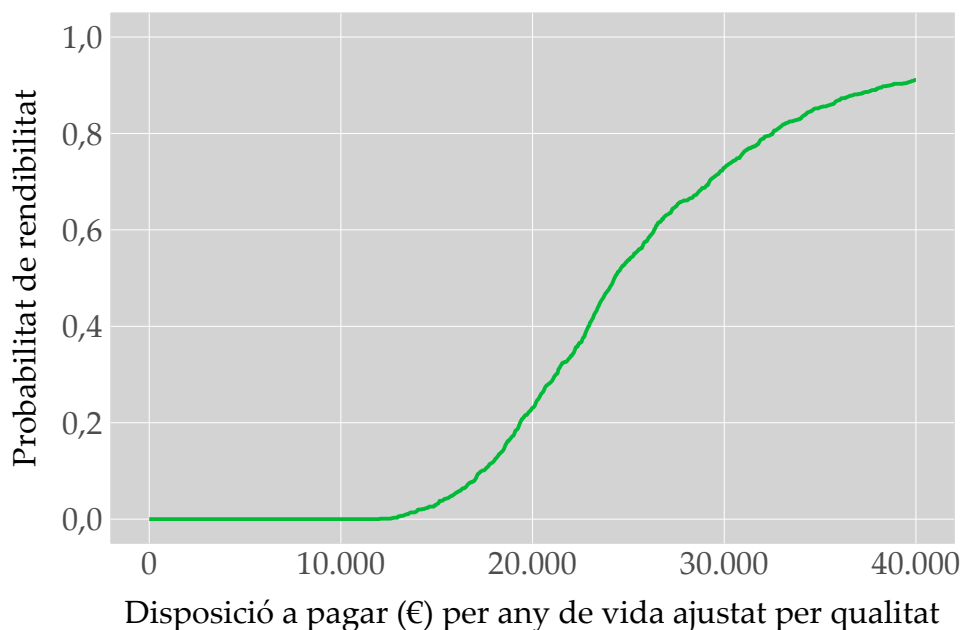


FIGURA 1.7: Corba d'acceptabilitat de cost-efectivitat per al programa de cribratge B respecte de l'estratègia de no cribratge

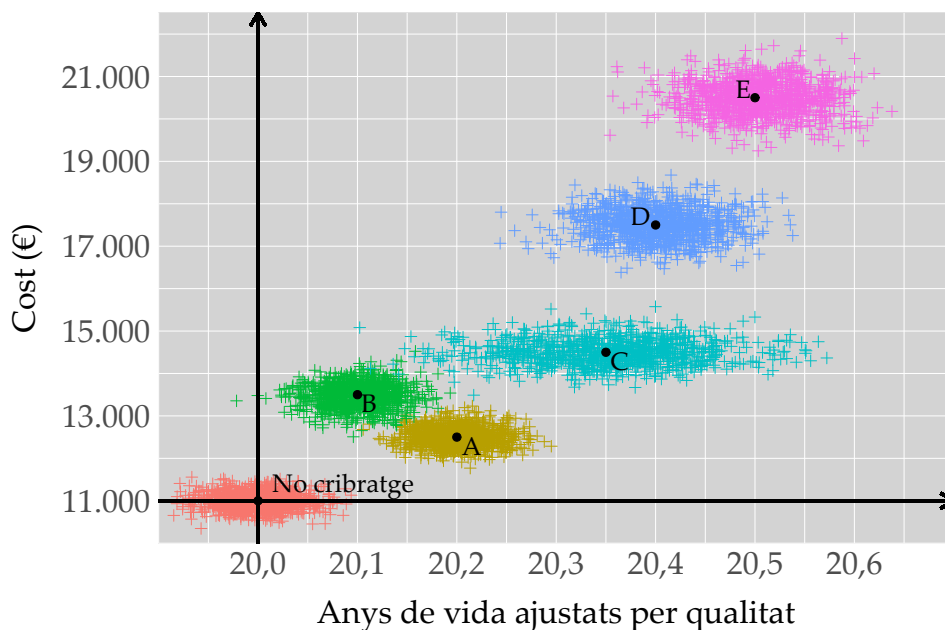


FIGURA 1.8: Diagrama de punts de l'anàlisi de sensibilitat probabilística de totes les alternatives de cribratge

sigui rendible es pot aproximar per la proporció de simulacions en les quals cadascuna assoleix el major benefici net esperat. Les corbes d'acceptabilitat de cost-efectivitat de les diferents alternatives de cribratge de la figura 1.8 es mostren en la figura 1.9. Per exemple, si el llindar és de 20.000 €, no hi ha cap incertesa associada a la decisió de rebutjar l'alternativa B basada en la

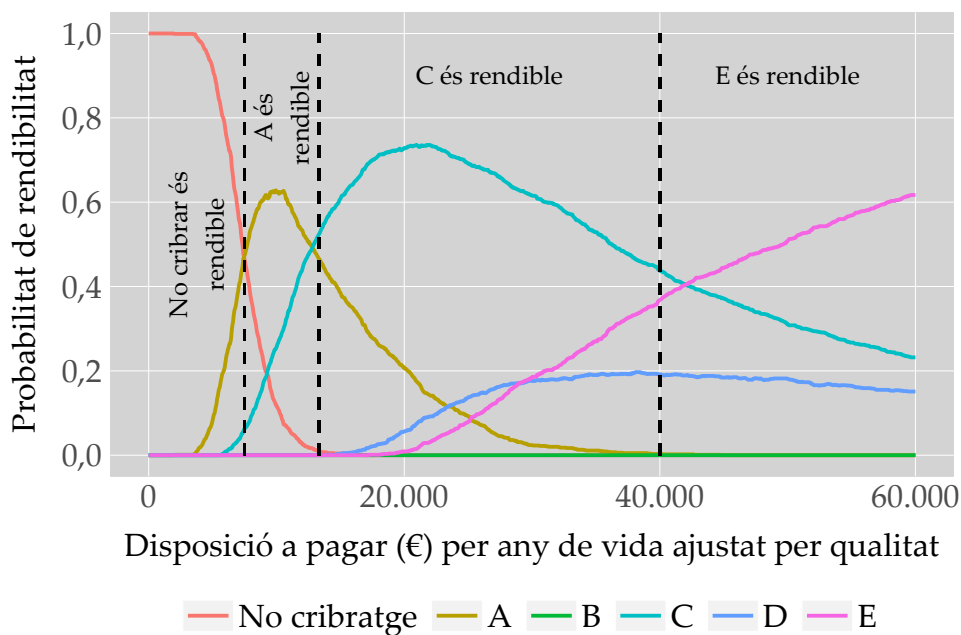


FIGURA 1.9: Corba d'acceptabilitat de cost-efectivitat per a totes les alternatives de cribratge

rendibilitat esperada, i pràcticament cap a rebutjar l'alternativa E. Si el llindar és de 30.000 €, la incertesa de la decisió és major: la probabilitat que el programa C proporcioni el major benefici net és 0,62, però és possible que altres alternatives siguin rendibles (els programes A, D i E, amb probabilitats 0,02, 0,18 i 0,18 respectivament).

És important tenir present que l'alternativa amb major probabilitat de ser rendible pot no ser l'alternativa amb major rendibilitat esperat i que proporciona un benefici net més alt [92]. Així, com es pot observar en la figura 1.9, entre els lindars 40.000 € i 42.050 €, s'espera que el programa E sigui més eficient —és a dir, produeix uns beneficis nets esperats més alts que el programa C—, però té una probabilitat menor que C de ser rendible. Aquest fet, que pot semblar contraintuïtiu, simplement és un reflex de l'asimetria de la distribució de les diferències en beneficis nets, de manera que la mitjana és major que la mediana. En altres paraules, quan el programa E és millor que el programa C, la diferència és major que quan el programa C és millor que l'E. Succeeix un fet anàleg entre els lindars 12.800 € i 13.333,33 €, en què s'espera que el programa A sigui el programa més eficient, tot i que amb una probabilitat menor que el programa C. En aquest cas, la mitjana és menor que la mediana. Per això, sempre cal representar quines de les alternatives s'espera que siguin rendibles juntament amb la seva probabilitat. En la figura 1.9 s'indica amb les línies discontinües la frontera d'acceptabilitat de cost-efectivitat.

Amb tot, la decisió d'incorporar una certa alternativa partint de la informació disponible sempre tindrà alguna possibilitat de ser errònia. Per exemple, segons la figura 1.9, el programa E té una probabilitat del 63% de no ser rendible quan els responsables de la decisió estan disposats a pagar 40.000 € per AVAQ. Així, el cost esperat de la incertesa depèn de la probabilitat que una decisió sigui errònia i de les conseqüències de prendre una decisió equivocada. Els costos previstos de la incertesa es poden interpretar com el *valor esperat de la informació perfecta* (VEIP), ja que una informació perfecta pot eliminar la possibilitat de prendre una decisió equivocada [93]. Com més gran sigui el VEIP, més gran és el cost d'oportunitat d'una decisió incorrecta. La informació és valuosa perquè redueix la possibilitat de prendre una decisió errònia i, per tant, redueix els costos de la incertesa associada a la decisió. El VEIP és també la suma màxima que el sistema sanitari hauria d'estar disposat a pagar per obtenir més evidència per informar d'una mateixa decisió en el futur [93]. Així, el VEIP es calcula habitualment en termes monetaris mitjançant l'ús del benefici monetari net, tot i que també es pot expressar mitjançant l'ús del benefici net en salut [86]. En el capítol 11 de Drummond *et al.* [7] es poden trobar els detalls de les alternatives metodològiques per al seu càlcul i de la seva aplicació per decidir si és rendible finançar més recerca per informar de futures decisions. A Claxton *et al.* (2001) [94] i Claxton *et al.* (2002) [93] es reporta un cas pràctic del càlcul i de l'aplicació del VEIP en l'avaluació d'un nou tractament farmacològic per a la malaltia d'Alzheimer.

1.6 Criteris de decisió

Un cop realitzada l'avaluació econòmica cal decidir quina alternativa escollir, per la qual cosa cal definir un criteri de decisió [36]. Si els recursos fossin il·limitats sempre s'escolliria l'alternativa que proporcionés una major efectivitat. Davant dels pressupostos limitats, també cal tenir en compte el preu per unitat addicional d'efectivitat que s'està disposat a pagar.

Continuant amb l'exemple de les alternatives de cribratge, sense limitació pressupostària el programa E seria l'escollit. Ara bé, si el decisor disposa d'una quantitat addicional màxima a pagar de 8.000€ per dona en el nou programa de cribratge, el programa E ja no seria una opció factible a oferir a tota la població (figura 1.10). Llavors, si es volen maximitzar els AVAQ es podrien oferir diferents programes a les dones. En concret, oferint el programa E a un 75% de la població i el programa C a la resta, els costos incrementals d'aplicar la nova estratègia de cribratge serien $1/4 \times 3.500€ + 3/4 \times 9.500€ = 8.000€$ i s'assoliria una efectivitat incremental d' $1/4 \times 0,35 + 3/4 \times 0,5 \approx 0,46$ AVAQ (punt vermell en la figura 1.10). Si per raons d'equitat no es volgués oferir alternatives diferents, l'alternativa òptima seria aquella amb major efectivitat amb un cost inferior al límit pressupostari. A l'exemple, l'alternativa òptima seria el programa D, que passaria a formar part de la frontera d'eficiència òptima un cop desestimat el programa E.

A part de considerar el límit pressupostari, per realitzar l'elecció cal determinar si l'alternativa amb major efectivitat la proporciona amb un cost incremental inferior al que s'està disposat a pagar. En l'exemple del cribratge, si

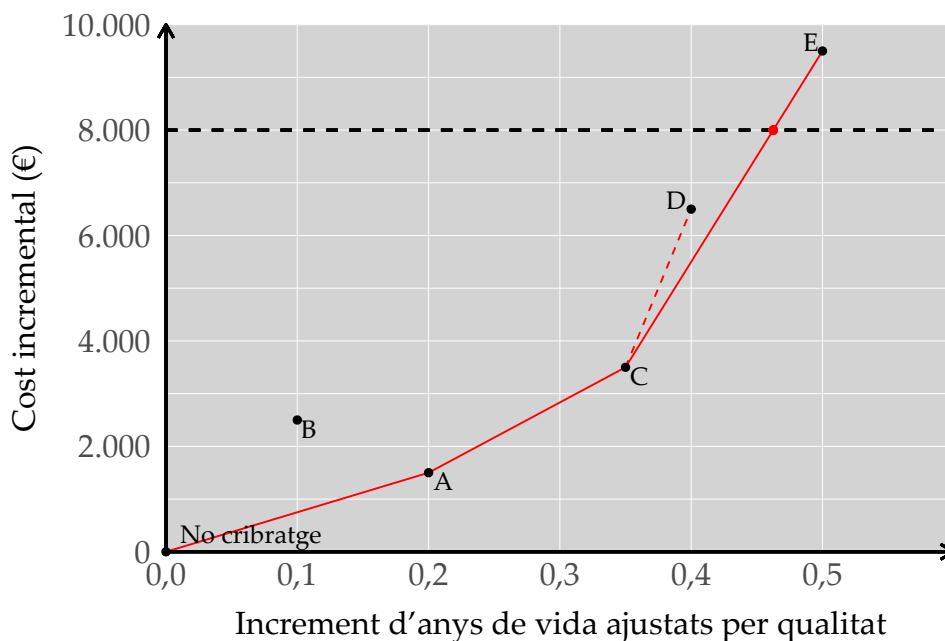


FIGURA 1.10: Exemple de presa de decisions en l'exemple de les alternatives de cribratge de càncer de mama

el llindar de rendibilitat fos de 35.000 € per AVAQ, el programa D en quedaria exclòs, atès que la seva ICUR és de 60.000 € per AVAQ en comparació amb el programa C, superior a la disposició a pagar màxima. Així, la següent alternativa més efectiva, el programa C, seria l'elecció òptima, ja que, a part de respectar el límit pressupostari, proporciona un increment d'efectivitat a un cost incremental de 13.333,33 €, que sí que es consideraria assumible. Per tant, per fer judicis sobre què és i què no és rendible cal fer-ho amb referència a un llindar de rendibilitat que representi els costos d'oportunitat [7, 95]. És inevitable fer una avaluació del llindar, perquè en cas de no ser explícit, el sol fet de prendre una decisió comporta un valor implícit pel llindar [7].

Per tal de fer comparacions entre intervencions en diferents problemes de salut, es poden construir *classificacions (league tables)* partint de la relació de cost-efectivitat. Aquestes classificacions permeten posar els resultats de les avaluacions econòmiques en un context més ampli, alhora que proporcionen una visió de l'eficiència de les diferents intervencions competint en la distribució dels recursos [36]. Tanmateix, l'ús de les classificacions partint de la relació de cost-efectivitat planteja diverses limitacions que susciten diverses crítiques, moltes de les quals associades a les característiques metodològiques pròpies de les anàlisis de cost-efectivitat. Així, els trets distintius de les anàlisis que més poden afectar la comparació de les relacions de cost-efectivitat són: la taxa de descompte; el mètode d'estimació de les utilitats en salut; l'interval de costos i de resultats utilitzat, i la selecció de les alternatives comparades [96]. Per tal que les classificacions siguin útils, els decisors haurien de poder avaluar la rellevància i la fiabilitat de l'evidència que proporciona l'avaluació econòmica en el seu propi entorn. A més a més, els autors de les avaluacions econòmiques han de reportar els mètodes i els resultats tan exhaustivament com sigui possible, atès que ajudarien molt en la construcció i l'ús adequat de les classificacions [96].

1.7 Mètodes estadístics en avaluació econòmica

Els mètodes estadístics són presents en tota avaluació econòmica, ja sigui basada en les dades d'un estudi clínic o en un model analític. En el cas d'avaluacions basades en dades individuals, els mètodes estadístics són necessaris, tal com ho són en l'anàlisi d'estudis clínics. Així, cal aplicar mètodes d'anàlisi de dades categòriques, variables discretes i variables contínues. Per tal d'establir mesures d'associació lliures de confusió, habitualment també cal fer servir mètodes de regressió, des de models lineals múltiples o models lineals generalitzats [97], amb especial protagonisme de la regressió logística [98], o models de dades categòriques o de recompte (*count data*) [99], fins a l'anàlisi de dades de supervivència [100-102]. Tanmateix, l'ús d'assajos clínics com a vehicle d'avaluacions econòmiques té una història més curta, per la qual cosa les propostes metodològiques específiques per a tractar els matisos de les anàlisis de cost-efectivitat s'han anat produint durant els darrers anys [7, 103].

Una de les preocupacions de qualsevol estudi és la mesura dels efectes causals de les exposicions o les intervencions. En els RCT, l'associació és de causalitat, atès que l'aleatorització assegura que els exposats i els no exposats poden ser intercanviables. Per contra, en estudis observacionals, l'associació no és generalment de causalitat, ja que els exposats i els no exposats no són generalment intercanviables [59]. Tanmateix, donat que el disseny observacional sovint és l'única alternativa factible, cal aplicar metodologies adequades que incorporin el context causal en l'anàlisi. Altrament, analitzar les dades únicament en termes estadístics, tot eliminant-ne el context causal, pot conduir a errors analítics i a interpretacions errònies de la realitat que s'intenta explicar [104].

Una de les particularitats de les dades econòmiques que requereixen nous mètodes o l'adaptació de les tècniques existents és l'asimetria en les dades de costos, que acostumen a presentar una cua llarga a la dreta de la distribució; és a dir, hi ha una proporció relativament petita de pacients amb valors relativament grans. A més a més, sovint una gran part dels pacients tenen cost zero perquè no consumeixen recursos rellevants, per la qual cosa s'han revisat i proposat diverses metodologies per tractar adequadament aquestes limitacions en les dades d'ús de recursos i costos [105]. Una altra limitació és la presència de dades mancants. El problema de la manca de dades també existeix en l'anàlisi de dades clíniques, però té algunes implicacions especialment importants per a l'anàlisi econòmica [7]. Tal com s'ha vist en la secció 1.2.5.1, l'estimació del cost es basa en dades d'ús de recursos, cadascun dels quals es multiplica per un cost unitari. Si falta algun d'aquests elements de recurs per a un pacient en particular, no és clar com calcular el cost total d'aquest pacient. Una situació similar existeix amb l'anàlisi de dades individuals sobre QVRS. Per identificar el millor mètode per tractar les dades mancants, cal avaluar els mecanismes que les ocasionen. Un altre aspecte que origina manca d'informació és el temps de seguiment, que sol diferir entre pacients, i que condueix a dades censurades. Existeixen mètodes per a l'anàlisi del temps fins a l'esdeveniment d'interès [100-102], que poden no ser adequades per analitzar dades de costos censurades i que, per tant, poden conduir a estimacions esbiaixades. Tots aquests aspectes, i d'altres, propis de l'anàlisi de dades per a l'avaluació econòmica i de les solucions metodològiques proposades, estan exhaustivament revisats en les publicacions de Willan i Briggs [103] i Glick *et al.* [106].

Una metodologia especialment adequada i útil és la modelització conjunta de dades longitudinals i temps fins a l'esdeveniment [107]. La modelització conjunta de dades longitudinals i de temps fins a l'esdeveniment és una àrea cada cop més productiva d'investigació estadística que avalua l'associació entre processos longitudinals i de supervivència. Millora la modelització longitudinal, ja que permet la inclusió de mecanismes d'abandonament no susceptibles de ser ignorats, alhora que millora els models de supervivència estàndards mitjançant la inclusió de covariables internes que depenen del temps. Els models conjunts d'efectes aleatoris compartits (*shared-effects joint models*) són un tipus de models conjunts en què els processos longitudinals i

de supervivència es connecten mitjançant un vector comú d'efectes aleatoris. Així doncs, no únicament són útils i adequats per obtenir estimacions no esbiaixades de la incidència o la mortalitat, sinó que també permeten estimar evolucions de marcadors longitudinals, que poden ser variables de costos. A més a més, la seva aplicació és factible en presència de dades mancants sense ocasionar la pèrdua de cap pacient en el procés d'inferència. Donada la relativa curta història del desenvolupament de programari estadístic que en permeti l'aplicació, no han estat gaire introduïts en anàlisis econòmiques, tot i que se'n poden trobar alguns exemples en la literatura [108-111].

Per altra banda, els mètodes estadístics d'aprenentatge automàtic (*machine learning*) o d'aprenentatge profund (*deep learning*) i la intel·ligència artificial han anat agafant protagonisme en la recerca biomèdica, sobretot arran de l'explosió de les *dades òmiques*, els estudis de descobriment de nous biomarcadors i el diagnòstic per imatge [112]. Per tant, també han anat introduint-se en avaluacions econòmiques, en què encara s'està avaluant la utilitat i les dificultats en l'aplicació d'algorismes d'aprenentatge automàtic i discutint sobre aquests aspectes [113, 114].

Com ja s'ha apuntat en la secció 1.3, les avaluacions econòmiques basades en models analítics procuren incorporar tota l'evidència disponible. De fet, és un dels seus avantatges, ja que quan hi ha diverses fonts d'evidència per a un paràmetre determinat, totes poden ser incorporades al model. Per tant, és habitual realitzar una revisió sistemàtica per identificar tota l'evidència vàlida disponible i posteriorment sintetitzar totes les estimacions mitjançant una metanàlisi per incloure-les en un model de decisió [7]. El més habitual en les metanàlisis és avaluar l'eficàcia i la seguretat de la nova tecnologia en comparació directa amb un comparador únic [115]. Ara bé, mitjançant, per exemple, el mètode de Bucher, es poden obtenir estimacions de comparacions indirectes quan la informació és limitada o inexistent [116]. Les metanàlisis en xarxa, o comparacions mixtes, permeten combinar estimacions de comparacions directes i indirectes, a fi d'augmentar la potència estadística [116]. A l'últim, en la secció 1.5 també s'ha apuntat la importància de les tècniques estadístiques per a l'avaluació de la incertesa en els models analítics.

1.8 Aplicació en problemes de salut

Hi ha un gran ventall d'exemples de problemes o situacions en les quals cal realitzar una decisió en salut i, per tant, una avaluació econòmica prèvia. La prevenció o el tractament de malalties amb majors taxes d'incidència i/o de mortalitat són probablement els que mereixen una especial atenció, per la major càrrega social que ocasionen.

El càncer de mama és el càncer més comunament diagnosticat entre les dones. Segons les últimes dades recollides pel Sistema Europeu d'Informació del Càncer, durant el 2018 es van diagnosticar un total de 32.825 nous casos de càncer de mama a Espanya; és, doncs, el tipus de tumor més freqüent entre les dones, per davant del càncer colorectal, d'úter, de pulmó i d'ovari,

i representa aproximadament el 30% dels càncers diagnosticats en les dones espanyoles [117]. Segons la mateixa font, durant el 2018 van morir 6.421 dones per càncer de mama, just per darrere del càncer colorectal (6.645) i per davant del càncer de pulmó (5.337), fet que representa aproximadament el 15% de morts per càncer en dones d'Espanya.

La malaltia cardiovascular és un dels problemes de salut més freqüents que afecten homes i dones a tot Europa i, cada cop més, als països de la Mediterrània oriental [118]. El 2015 hi va haver aproximadament 11 milions de nous casos als països membres de la Societat Europea de Cardiologia; se sobrepassaren, així, els 83 milions de casos prevalents [118]. L'impacte social de la malaltia cardiovascular és enorme, tant pels costos directes associats als recursos assistencials com pels costos indirectes relacionats amb l'absentisme, la productivitat perduda i la mortalitat prematura evitable [119-121]. Les estadístiques de mortalitat mostren que la malaltia cardiovascular continua sent la causa més freqüent de mort a Europa, ja que representa el 45% de totes les morts —el 49% de les morts entre les dones i el 40% entre els homes— [118, 122, 123], tot i que amb grans desigualtats entre els països amb ingressos mitjans, amb més del 50% de les morts, i els països amb ingressos alts de l'Europa occidental, on la mortalitat cardiovascular representa menys del 30% de totes les morts [118]. Dels 4 milions de persones que moren anualment de malalties cardiovasculars a Europa, 1,4 milions es produeixen abans dels 75 anys [122, 123]. La síndrome coronària aguda (SCA) és un subgrup de malaltia cardiovascular que inclou l'angina inestable i l'infart de miocardi, que són els esdeveniments cardiovasculars amb major cost associat, sobretot durant els primers anys [124, 125].

1.8.1 Cribratge poblacional de càncer de mama

Atès que la majoria de factors de risc coneguts del càncer de mama no són modificables, no existeixen programes poblacionals de prevenció primària. En conseqüència, la detecció precoç és una prioritat entre els programes de salut pública, amb l'objectiu de millorar el pronòstic de la malaltia i reduir-ne la mortalitat. Juntament amb el desenvolupament de noves teràpies adjuvants, la detecció precoç del càncer de mama en redueix la mortalitat i pot millorar la qualitat de vida de les dones diagnosticades precoçment mitjançant exàmens mamogràfics [126, 127]. No obstant això, el cribratge poblacional és costós i pot causar danys —per exemple, falsos positius, sobrediagnòstic— a moltes d'elles [128-131]. La figura 1.11 mostra els danys i els beneficis associats al cribratge per mamografia. Per tant, la justificació dels programes de cribratge poblacional de càncer de mama ha d'implicar el fet que els beneficis globals superin els danys a un cost raonable. A més a més, una avaluació econòmica és especialment necessària quan el cribratge està finançat per recursos públics.

Els programes de cribratge organitzats per a la detecció precoç del càncer de mama ofereixen per igual proves de cribratge en totes les dones elegibles,

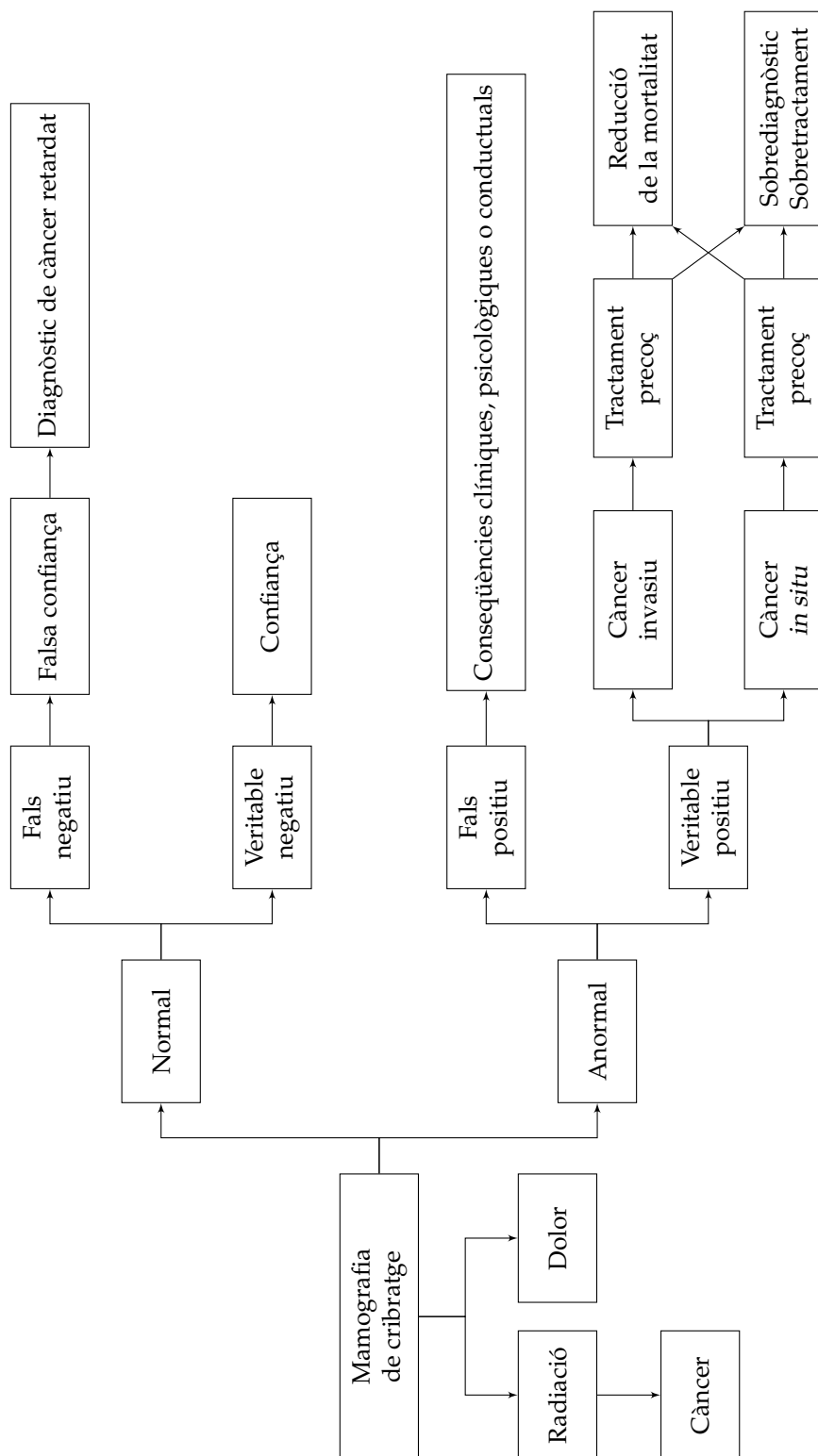


FIGURA 1.11: Danys i beneficis del cribratge mamogràfic

Adaptat d'Armstrong et al. «Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians». *Annals of Internal Medicine*, 2007, vol. 146, p. 516-526.

independentment del seu risc individual. Per exemple, les directrius europees recomanen oferir un cribratge per mamografia a dones de 50 a 69 anys cada dos anys [132]. Aquest paradigma uniforme o de “cafè per a totes” ha anat derivant cap a la personalització d’estratègies de detecció basades en el risc de càncer de mama. En aquest sentit, l’Institute of Medicine va identificar el cribratge personalitzat com a crucial per millorar la detecció precoç del càncer de mama [133]. Ara bé, disposar d’una estimació fiable del risc individual de càncer de mama és fonamental per desenvolupar programes de cribratge personalitzats i optimitzar l’ús dels recursos en una població. En general, es considera que una eina de predicció ha de tenir una capacitat de discriminació mínima, amb una àrea sota la corba ROC superior a 0,7 [98]. Malauradament, l’aplicació en població espanyola dels models de predicció de risc de càncer de mama ha mostrat una capacitat de discriminació limitada [134].

1.8.2 Prevenció cardiovascular

Qualsevol intervenció en població general destinada a reduir els factors de risc cardiovasculars estalvia costos al sistema sanitari i millora la salut de la població, tal com s’ha evidenciat en diversos estudis [135-138]. Tanmateix, hi ha poblacions d’alt risc cardiovascular amb factors de risc no modificables, com els pacients amb malaltia renal crònica (MRC), als quals és necessari aplicar altres intervencions. L’MRC afecta aproximadament un 7,2% de persones majors de 30 anys, però la prevalença augmenta fins al 23-36% en població major de 65 anys [139], i s’espera que augmenti en les properes dècades [140]. Tenint en compte aquestes tendències, i donat que els pacients amb MRC tenen major risc d’esdeveniment cardiovascular, independentment de la presència d’altres comorbiditats de risc, com la diabetis i la hipertensió [141, 142], està justificada la inversió de recursos en la recerca de nous marcadors que ajudin a millorar l’estimació de risc cardiovascular en aquesta població, subestimat pels models de risc estàndard [143]. A més a més, nous biomarcadors del risc cardiovascular podrien obrir la porta a proposar i desenvolupar nous tractaments.

A banda, també cal avaluar les intervencions disponibles per millorar els resultats en salut a curt i llarg termini després de patir un esdeveniment cardiovascular. Un exemple és la intervenció coronària percutània (ICP) en cas de patir un esdeveniment d’SCA. Entre el 17 i el 34% dels pacients ingressats per SCA són ancians [144, 145] i segons l’evidència actual el seu tractament hauria de ser similar al dels pacients més joves. En pacients amb infart de miocardi amb elevació del segment ST (STEMI, de l’anglès *ST-segment elevation myocardial infarction*), l’estàndard assistencial és la reperfusió immediata [146]. La ICP és l’estratègia de reperfusió preferida, per davant de la trombòlisi, sense límit d’edat superior. En pacients amb SCA sense elevació del segment ST (NSTEMI-ACS, de l’anglès *non-ST-segment elevation acute coronary syndrome*) es recomana seguir una estratègia invasiva si els pacients tenen un risc elevat [147], la qual cosa representa aproximadament 2/3 dels

pacients amb NSTEMI-ACS. L'estratègia invasiva inclou una angiografia coronària urgent seguida de revascularització, que sol realitzar-se amb una ICP. Per a pacients de risc elevat, l'estàndard d'atenció és una estratègia invasiva rutinària. En els pacients més grans, l'estratègia invasiva rutinària redueix l'ocurrència d'infart agut de miocardi i de revascularització posterior, comparada amb l'estratègia selectiva [148]. Tot i que les directrius actuals recomanen una estratègia invasiva en tots els pacients amb STEMI i en la majoria de pacients amb NSTEMI-ACS, registres poblacionals han descrit un ús més baix de la ICP i l'angiografia en pacients ancians amb SCA [149, 150], possiblement associat a la presència de comorbiditats i a l'aparició de símptomes atípics [151], així com un major risc de complicacions de la ICP [144, 152-155]. D'altra banda, també és possible que la infrautilització de la ICP estigui en part causada per la manca d'estimacions de la relació de cost-efectivitat del seu ús en pacients ancians amb SCA. Per aportar major evidència, es va dur a terme el projecte EUROTRACS [156], que es va centrar a examinar l'efectivitat de diferents intervencions per millorar els resultats de malalties cardiovasculars a Europa.

Capítol 2

Hipòtesis i objectius

2.1 Hipòtesi general

La modelització matemàtica i estadística proporciona informació als decisors, clínics i gestors, per prendre una decisió, ja que constitueix la metodologia necessària i vàlida per aportar evidència, ja sigui mitjançant la simulació de possibles escenaris alternatius o amb l'anàlisi de dades reals.

2.2 Hipòtesis específiques

Sobre les estratègies de cribratge poblacional de càncer de mama:

- H1 La modelització de la història natural de la malaltia permetrà simular escenaris hipotètics de diverses estratègies de cribratge poblacional basat en el risc, i comparar-les amb el cribratge uniforme actual.
- H2 Incloure mesures longitudinals de densitat mamària en models conjunts de dades longitudinals i temps fins a un esdeveniment permetrà obtenir mesures de risc de càncer de mama més acurades.

Quant a la prevenció cardiovascular:

- H3 La inclusió de nous biomarcadors d'ateromatosis subclínica en l'estratificació del risc cardiovascular podria millorar la capacitat predictiva en població amb malaltia renal crònica.
- H4 Estendre l'ús de la intervenció coronària percutània a pacients ancians hospitalitzats amb un esdeveniment de síndrome coronària aguda pot proporcionar beneficis en salut a uns nivells d'eficiència assumibles per part del finançador.

2.3 Objectius

L'objectiu general és millorar la presa de decisions sanitàries, tant en el vessant de la pràctica clínica com en el de la gestió, mitjançant la utilització de models matemàtics i estadístics. Els objectius específics són:

- O1 Avaluar i comparar la rendibilitat de diferents estratègies de cribratge poblacional de càncer de mama mitjançant un model probabilístic.
- O2 Desenvolupar mesures del risc del càncer de mama mitjançant models conjunts de dades longitudinals i temps fins a l'esdeveniment.
- O3 Determinar l'associació i la capacitat predictiva de nous biomarcadors d'esdeveniment cardiovascular en pacients amb malaltia renal crònica.
- O4 Avaluar l'eficiència de l'augment de l'ús de la intervenció coronària percutània per a la gestió de pacients de 75 anys o més hospitalitzats amb un esdeveniment de síndrome coronària aguda mitjançant un model probabilístic.

Capítol 3

Mètodes i resultats

Els materials i mètodes utilitzats i els resultats obtinguts que donen resposta als objectius de la tesi es presenten en les seccions següents. La taula 3.1 relaciona els diferents objectius plantejats en la tesi i les seccions on es pot trobar el detall de la metodologia i dels resultats.

TAULA 3.1: Relació entre els objectius de la tesi i els capítols on es detalla la metodologia i es presenten els resultats

Objectiu	Secció
O1. Avaluar i comparar el cost-efectivitat de diferents estratègies de cribratge poblacional de càncer de mama mitjançant un model probabilístic	Secció 3.1
O2. Desenvolupar mesures del risc del càncer de mama mitjançant models conjunts de dades longitudinals i temps fins a l'esdeveniment	Secció 3.2
O3. Determinar l'associació i la capacitat predictiva de nous biomarcadors d'esdeveniment cardiovascular en pacients amb malaltia renal crònica	Secció 3.3
O4. Avaluar l'eficiència de l'augment en l'ús de la intervenció coronària percutània per a la gestió de pacients ancians hospitalitzats amb un esdeveniment de síndrome coronària aguda mitjançant un model probabilístic	Secció 3.4

3.1 Mètodes i resultats de l'objectiu 1

En aquesta secció es presenten els mètodes utilitzats i els resultats obtinguts per donar resposta al primer dels objectius plantejats en aquesta tesi. Tota la metodologia i els resultats estan continguts en el manuscrit original Vilapri-nyo *et al.* (PLOS ONE, 2014), la informació del qual es detalla a continuació, junt amb el resum traduït al català.

Títol: Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer

Autors: Vilapriño E, Forné C, Carles M, Sala M, Pla R, Castells X, Domingo L, Rue M; Interval Cancer (INCA) Study Group

Revista: PLOS One. 2014 Feb 3;9(2):e86858

Factor d'impacte de l'any de la publicació: 3,234 (percentil 85,1 a *Multidisciplinary Sciences*)

DOI: 10.1371/journal.pone.0086858

Resum:

El paradigma d'una mateixa estratègia única i uniforme per a totes les dones en el cribratge organitzat de càncer de mama està canviant cap a un enfocament personalitzat. El present estudi té dos objectius: (1) realitzar una avaluació econòmica i avaluar les relacions dany-benefici de les estratègies de cribratge que varien en la seva intensitat i en els intervals d'edat basant-se en el risc de càncer de mama, i (2) estimar el guany en termes de reducció de costos i de danys del cribratge basat en risc respecte a la pràctica habitual. Hem utilitzat un model probabilístic i dades de registres i programes de cribratge en població espanyola, així com estudis clínics, per estimar el benefici, el dany i els costos al llarg del temps de 2.624 estratègies de cribratge, uniformes o basades en el risc. Es van definir quatre grups de risc —baix, moderat-baix, moderat-alt i alt— basats en la densitat mamària, els antecedents familiars de càncer de mama i els antecedents personals de biòpsia de mama. Les estratègies basades en el risc es van obtenir combinant la periodicitat de l'examen (anual, biennal, triennal i quinquennal), les edats inicials (40, 45 i 50 anys) i les edats finals (69 i 74 anys) en els quatre grups de risc. S'han calculat relacions de cost-efectivitat incremental i de dany-benefici incremental per seleccionar les estratègies òptimes. En comparació amb les estratègies basades en el risc, de les uniformes en resulta un benefici molt menor per a un cost específic. S'han obtingut reduccions properes al 10% en costos i superiors al 20% en resultats de falsos positius i casos sobrediagnòstics amb estratègies basades en el risc. El cribratge òptim es caracteritza per periodicitats quinquennals o triennals per als grups de risc baix o moderats, i periodicitat anual per al grup d'alt risc. Les estratègies basades en el risc poden reduir-ne els danys i els costos. Cal desenvolupar mesures precises

del risc individual i treballar en la implementació d'estratègies de cribratge basades en el risc.

Cost-Effectiveness and Harm-Benefit Analyses of Risk-Based Screening Strategies for Breast Cancer

Ester VilaprinYO^{1,2,9}, Carles Forné^{1,2,9}, Misericordia Carles³, Maria Sala^{4,5}, Roger Pla^{6,7}, Xavier Castells^{4,5}, Laia Domingo⁴, Montserrat Rue^{1,2,5*}, the Interval Cancer (INCA) Study Group[¶]

1 Basic Medical Sciences Department, Biomedical Research Institut of Lleida (IRBLLEIDA), Lleida, Catalonia, Spain, **2** Basic Medical Sciences Department, University of Lleida, Lleida, Catalonia, Spain, **3** Economics Department and Research Centre on Industrial and Public Economics (CREIP), Rovira i Virgili University, Reus, Catalonia, Spain, **4** Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar, Mar Teaching Hospital, Barcelona, Catalonia, Spain, **5** Health Services Research Network in Chronic Diseases (REDISSEC), Spain, **6** Surgery Department, Rovira i Virgili University, Reus, Catalonia, Spain, **7** General and Digestive Surgery Department, Joan XXIII Teaching Hospital, Tarragona, Catalonia, Spain

Abstract

The one-size-fits-all paradigm in organized screening of breast cancer is shifting towards a personalized approach. The present study has two objectives: 1) To perform an economic evaluation and to assess the harm-benefit ratios of screening strategies that vary in their intensity and interval ages based on breast cancer risk; and 2) To estimate the gain in terms of cost and harm reductions using risk-based screening with respect to the usual practice. We used a probabilistic model and input data from Spanish population registries and screening programs, as well as from clinical studies, to estimate the benefit, harm, and costs over time of 2,624 screening strategies, uniform or risk-based. We defined four risk groups, low, moderate-low, moderate-high and high, based on breast density, family history of breast cancer and personal history of breast biopsy. The risk-based strategies were obtained combining the exam periodicity (annual, biennial, triennial and quinquennial), the starting ages (40, 45 and 50 years) and the ending ages (69 and 74 years) in the four risk groups. Incremental cost-effectiveness and harm-benefit ratios were used to select the optimal strategies. Compared to risk-based strategies, the uniform ones result in a much lower benefit for a specific cost. Reductions close to 10% in costs and higher than 20% in false-positive results and overdiagnosed cases were obtained for risk-based strategies. Optimal screening is characterized by quinquennial or triennial periodicities for the low or moderate risk-groups and annual periodicity for the high-risk group. Risk-based strategies can reduce harm and costs. It is necessary to develop accurate measures of individual risk and to work on how to implement risk-based screening strategies.

Citation: VilaprinYO E, Forné C, Carles M, Sala M, Pla R, et al. (2014) Cost-Effectiveness and Harm-Benefit Analyses of Risk-Based Screening Strategies for Breast Cancer. PLoS ONE 9(2): e86858. doi:10.1371/journal.pone.0086858

Editor: Anna Sapino, University of Torino, Italy

Received: April 8, 2013; **Accepted:** December 16, 2013; **Published:** February 3, 2014

Copyright: © 2014 VilaprinYO et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by grants PS09/01340 and PS09/01153 from the Health Research Fund (Fondo de Investigación Sanitaria) of the Spanish Ministry of Health. The authors thank the Breast Cancer Surveillance Consortium and the funding that the BCSC received from the National Cancer Institute (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, and HHSN261201100031C). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: montse.rue@cmb.udl.cat

¶ These authors contributed equally to this work.

¶ Membership of the Interval Cancer (INCA) Study Group is provided in the Acknowledgments.

Introduction

Early detection of breast cancer (BC) reduces mortality and may improve quality of life for most of the women diagnosed early by mammographic exams [1]. Nevertheless, screening healthy women is expensive and may cause harms (e.g. false positive results, overdiagnosis) in many of them [2–5]. In order for organized screening programs to be justified in this time of economic constraints, overall benefits should outweigh harms at a reasonable cost. Moreover, an economic evaluation is especially necessary when screening is funded by community resources.

Organized screening programs for early detection of BC provide screening services where all eligible women are treated as equal risk. For instance, the European guidelines recommend offering mammography screening to women aged 50–69 every two years [6]. This one-size-fits-all or uniform paradigm is starting to shift toward personalizing screening strategies based on breast

cancer risk. In 2005 the Institute of Medicine (IOM) identified that personalized screening was crucial to improving the early detection of breast cancer [7]. More recently, Schousboe *et al.* [8], using a Markov microsimulation model, found that the cost-effectiveness of screening mammography depended on a woman's age, breast density, family history, and history of breast biopsy. Based on their results, mammography every two years was cost-effective for women aged 40 to 49 years with relatively high breast density or additional risk factors for breast cancer. And mammography every three to four years was cost-effective for women aged 50 to 79 years with low breast density and no other risk factors. van Ravesteyn *et al.* [9], using different microsimulation models, determined that women aged 40 to 49 years with a twofold increase in risk have similar harm-benefit ratios for biennial screening mammography as average-risk women aged 50 to 74 years.

In a previous study, we performed an economic evaluation of uniform screening strategies that had different periodicities and varied in the ages at starting or ending the screening exams [10]. The present study has two objectives that extend our previous work: 1) To perform an economic evaluation and to assess the harm-benefit ratios of screening strategies that vary in their intensity and interval ages based on BC risk; and 2) To estimate the gain in terms of cost and harm reductions using risk-based screening with respect to the usual practice.

Methods

The model and model inputs

We used the probabilistic model developed by Lee and Zelen (LZ), which has been described elsewhere [11–13]. Further details of the model can be found in Appendix S1, section A. The model assumes a four-state progressive disease with S_0 : disease-free state, S_p : preclinical state (asymptomatic disease that can be diagnosed by a special exam), S_c : clinical state (diagnosis by symptomatic detection), and S_d : death from BC. The LZ model consists of a set of equations that allow to estimate the cumulative probability of death for a particular cohort exposed to a specific screening scenario or to no screening, after T years of follow-up. Since the model is analytical, for each specific set of inputs, the model run produces the same results. The model also provides incidence and prevalence of BC over time, both measures necessary for the estimation of treatment and follow-up costs.

The model requires input data that was obtained from different sources. BC incidence and survival, and mortality from other causes refer to cohorts born in Catalonia (Spain) in the period 1948–1952 [10,14–17]. The sojourn time in the pre-clinical state, the distribution of stages at diagnosis (Table S1 in Appendix S1) and the sensitivity of mammography were obtained by Lee and Zelen from published randomized clinical trials and observational studies [12]. Based on previous work of Zelen and Feinleib [18] and Day and Walter [19] on the randomized clinical trial of the Health Insurance Plan (HIP), it was assumed that the preclinical sojourn time follows an exponential distribution with an age dependent mean equal to 2 years for age < 40 and 4 years for age > 50. In the (40–50] age interval the mean sojourn time increases linearly from 2 to 4 years. The additional inputs are described below in the next subsections. All the calculations assumed an initial population of 100,000 women at birth. The time horizon for the study was 40–79 years of age.

The research protocol was approved by the institutional review board and ethics committee of the Hospital Universitari Arnau de Vilanova de Lleida (Spain) which waived the need for informed consent.

Risk of invasive breast cancer

We started estimating the age-specific risk of invasive BC for our study cohort, using the model published elsewhere by Martinez-Alonso *et al.* [17]. Details of the model can be found in Appendix S1, Section B.1. Then, following Tice *et al.* [20] and Schousboe *et al.* [8], age-specific BC risk groups were defined according to the following variables: breast density (measured using the Breast Imaging Report and Database System (BI-RADS) categories 1 to 4 [21]), family history of BC in first degree relatives (yes/no) and personal history of breast biopsy (yes/no). Details can be found in Appendix S1, section B.2.

We obtained four aggregated risk groups that combined the profiles of women that had similar levels of BC incidence over time: 1) Low (L) risk which included Category 1 breast density with at most one risk factor - family history or breast biopsy - and

Category 2 breast density with no risk factors; 2) Medium-Low (ML) risk which included Category 1 breast density with two risk factors, Category 2 breast density with one risk factor, and Categories 3 or 4 breast density with no risk factors; 3) Medium-High (MH) risk which included Category 2 breast density with two risk factors, Categories 3 or 4 breast density with one risk factor; and 4) High (H) risk which included Categories 3 or 4 breast density with two risk factors. The frequency distributions of the risk groups was 39.6%, 42.8%, 15.6% and 2.0% for L, ML, MH and H, respectively.

The incidence rates of the four aggregated risk groups were estimated as weighted sums of detailed incidence curves (see Section B.2, Tables S2, S3, and Figure S1 in Appendix S1). The weights were based on the prevalences of each combination of risk factors obtained from the Risk Estimation Dataset of the Breast Cancer Surveillance Consortium (BCSC) [22].

The screening strategies

We analyzed 2,625 screening strategies, 24 of them uniform and 2,601 risk-based. The risk-based strategies were obtained combining the exam periodicity (annual (A), biennial (B), triennial (T), and quinquennial (Q, [every five years])), the starting ages (40, 45 and 50 years) and the ending ages (69 and 74 years) in the four risk groups, L, ML, MH and H. In the following sections, uniform strategies are abbreviated as B5069 or B4574, for biennial exams in the 50–69 or in the 45–74 age groups, respectively. Risk-based strategies are abbreviated with four strings, e.g. Q5069-Q4574-T4574-A4074, that correspond to the L, ML, MH and H risk groups, respectively. A sample of the studied screening strategies is presented in Table S4 in Appendix S1.

The benefits

For each screening strategy and for the background, we measured the benefit of screening with two outcomes: the number of lives extended, LE, and the number of quality-adjusted life years gained, QALY. Because of the lack of Spanish data, the QALYs were estimated using the work of Lidgren *et al.* [23] in a sample of 361 Swedish women with localized, recurrent, or metastatic breast cancer (See Table S5 in Appendix S1). We considered the Lidgren's study more robust and suitable than other studies that used expert opinion or healthy population to obtain quality of life estimates associated with breast cancer. We used the values obtained from the EuroQol EQ-5D in the Lidgren's study. For women that did not die of BC we considered a loss of QALYs in the first five years following the diagnosis. For women that died of breast cancer, we considered that the last four years of their lives or the time from diagnosis to death, if they lived less than four years, were spent in a metastatic stage, independently of the stage at diagnosis. See section C in Appendix S1 for further details.

The harms

False positive (FP) results. We used the FP rates for non-invasive and invasive tests obtained from the Cumulative Risk of False Positive Study (RAFP) study which included 74 distinct radiology units in eight regions of Spain, from March 1990 to December 2006 [2]. The RAFP study included 1,565,364 women that underwent 4,739,498 mammographic exams. The FP rates were age and exam specific. We multiplied the FP rates by the number of women at risk for BC, at each specific exam, to estimate the number of women that would receive additional non-invasive (e.g. ultrasound) or invasive tests (e.g. biopsy). See further details in Appendix S1, section D and Tables S6 and S7.

Interval cancers and false-negative (FN) results. In our model, true interval tumors correspond to those that appear

between exams and were not in the pre-clinical state when the previous exam was performed. FN cases are tumors that were not detected in the previous exams due to lack of sensitivity of the screening test. We considered that all tumors in pre-clinical state in the previous exam were FN.

Overdiagnosis. Screening may cause overdiagnosis when it detects tumors which would never have been diagnosed during a lifetime without screening because of the lack of progressive potential or death from other causes. To estimate overdiagnosis we made some additional assumptions. We differentiated between overdiagnosis of invasive BC and ductal carcinoma in situ (DCIS). For both types of tumors we assumed that: 1) overdiagnosis only happens when a mammographic exam is performed, 2) a woman with an overdiagnosed tumor would not die of breast cancer, and 3) QALYs and costs of treatment (initial and follow-up) for women with overdiagnosed tumors are the same as for Stage I BC.

Overdiagnosis of invasive breast cancer. Estimates of overdiagnosis show high variation depending on the study design and the method used [3,4,17,24–28]. Based on the reported data, an overdiagnosis rate of 15% can be considered a sensible value.

Using the incidence model described in Appendix S1, section B.1 [17], and taking into account the distribution of the sojourn times in the preclinical state and the sensitivity of mammography (as in the LZ model), we estimated the number of BC cases that would be detected in the screening exams. Then, for each screening strategy, we assumed an overdiagnosis rate of 15% in the mammography exams. This assumption makes it possible to associate overdiagnosis with mammography exams, in the sense that more intensive screening strategies are considered to produce a higher overdiagnosis burden. For any screening strategy, an overdiagnosis estimate of 15% of the screen-detected cases gives an overall estimate lower than 15%, depending on the distribution of exam-detected and interval cases. For further details about how the overdiagnosis rate has been applied see section E and Table S8 in Appendix S1.

DCIS attributable to screening. To estimate the impact of screening on detection of DCIS we obtained the incidence and Census data from the Girona and Tarragona Cancer Registries in the period 1983–2008. Data on mammography use was obtained, for the Girona and Tarragona provinces, from three health surveys performed in the years 1994, 2002 and 2006 [29,30]. Section F in Appendix S1 explains in detail the analysis conducted to estimate the excess of DCIS attributable to screening. From this analysis we estimated an excess of 31.13 DCIS cases per 100,000 mammograms, with respect to a strategy of no screening (Table S9 and Figures S2 and S3 in Appendix S1).

Because DCIS is treated when detected, it is not possible to accurately estimate the fraction of detected DCIS that would progress to invasive disease. A review of the literature showed that between 14% and 53% of DCIS may progress to invasive cancer over a period of 10 or more years [31]. In our study we have assumed that 1/3 of the DCIS detected by mammography would progress to invasive cancer. With this assumption, the estimated excess number of DCIS attributable to screening was approximately $21 \approx 2/3 * 31.13 = 20.75$ per 100,000 mammograms, or 0.21 per 1,000 mammograms. In the sensitivity analysis we have estimated the proportion of DCIS that progress to be equal to 2/3 or to 1/6 of 31.13 per 100,000 mammograms.

Costs

We have adopted the perspective of the national health system and considered only direct healthcare costs. We have partitioned the estimation of costs into four parts: screening and diagnosis confirmation, initial treatment, follow-up and advanced care costs.

All costs were valued in 2012 euros and both costs and outcomes have been discounted at an annual rate of 3%, according to the economic evaluation guidelines of the Spanish Ministry of Health [32].

The costs of screening mammograms, complementary tests and administrative expenses were obtained from the Early Detection Program of *Parc de Salut Mar* (PSMAR) in the city of Barcelona. Data on treatment costs were obtained from a database that included 592 women consecutively diagnosed and initially treated for BC at the PSMAR in Barcelona in the period January 1st, 2000–December 31, 2003 [10].

Cost-effectiveness and harm-benefit analyses

To compare the relative costs and outcomes of the different strategies, we calculated the incremental cost-effectiveness ratio (ICER). The ICER is defined as the ratio of the change in costs to the change in effects of a specific intervention compared to an alternative. The ICER indicates the additional cost of obtaining one additional unit of outcome. We obtained the cost-effectiveness frontier, also called the Pareto frontier, which contains the efficient alternatives for which no alternative policy exists that results in better effects for lower costs.

To perform a harm-benefit analyses, we ordered the studied strategies from less to more adverse effects and obtained the incremental harm-benefit ratio of each strategy in relation to the previous one. We also obtained the harm-benefit frontier.

Selection of optimal strategies

To search for optimal strategies taking into account benefit, costs and harms, we selected the most recommended uniform strategy in Europe, biennial exams in the 50–69 age interval (B5069), or the alternative towards which some countries are moving, biennial exams in the 45–74 age interval (B4574), as reference strategies. Then, for each reference strategy we obtained the intersection of the subsets that contained strategies with similar benefit (between 1 and 1.05 times) than the reference strategy and lower cost and harms in terms of FP results and overdiagnosed cases (invasive and DCIS). The resulting strategies were located at or near the cost-effectiveness and harm-benefit frontiers with values in the x-axis near the B5069 or B4574 benefit values. We did not include the FN results in the intersection but we assessed them in the resulting optimal subset.

Validation of the model

We have compared our results with the results of three published reviews, the Cochrane systematic review [33], the Independent UK Panel on Breast Cancer Screening review [34], and the Euroscreen comprehensive review of European screening programs [35]. In addition, we have checked the results of the INterval CAncer (INCA) study in Spain, which included 645,764 women aged 45/50 to 69 years that participated biennially in seven population-based screening programs, from January 2000 to December 2006 (not yet published). A total of 1,508,584 mammograms were included in the study. The cohort was followed until June 2009 for breast cancer identification, resulting in 5,311 cases screen-diagnosed and 1,682 interval cancers.

We have compared the following summary indicators in the INCA study and the uniform B4569 strategy of our model: 1) frequencies of screen-detected and interval cancer, by age-group, 2) sensitivity of the program defined as the ratio of the number of tumors detected in the screening exams between all the detected tumors, 3) distribution of true interval cases and FN, by time since last mammogram, and 4) distribution of stages at diagnosis, by type of detection (screening or symptomatic).

Sensitivity analysis

There is uncertainty associated with the model inputs and there is also uncertainty associated with the model structure. It is complex and computationally intensive to obtain the variance of the model estimates. Instead, we performed univariate sensitivity analyses to study the impact on our conclusions when some of the inputs were modified. First, we changed the four risk group distributions assuming that 20% of women in the L, ML, and MH groups migrated to the next higher risk group. The new risk group distributions were 31.7%, 42.1%, 21.1% and 5.1%, for L, ML, MH and H, respectively. Second, we changed the amount of overdiagnosis of invasive tumors to 0%, 5% and 25%. Third, we changed the excess of DCIS to 0.1 and 0.26 per 1,000 mammograms. Fourth, we tested the effect of changing the costs of cancer treatment to two-fold and five-fold the costs of the main analysis. Fifth, we assessed the effect of changes in the disutility by false-positive result on QALY. We used zero and two times the disutility of the main analysis.

Data availability

All the input data will be available to researchers upon request.

Results

Cost-effectiveness and harm-benefit analyses

Benefits, harms, and costs of each screening strategy were obtained as a function of the risk-groups' incidence and the screening characteristics (periodicity and age-interval of exams by risk group). Figures 1 and 2 contain an overview of benefits, harms, and costs of all 2,625 strategies evaluated. The strategies that gave the best value for money can be found in Tables S10 and S11 in Appendix S1.

Measuring effectiveness with LE. Figure 1 and Table S10 in Appendix S1 present the results of the cost-effectiveness and harm-benefit analyses. Table 1 (A) shows two selected strategies that improve on the B5069 uniform strategy and two that improve on the B4574 uniform strategy. As an example, compared to B5069, the optimal strategy Q5074-Q5074-T5074-A5074 for the L, ML, MH and H risk groups, respectively, has 3.8% higher benefit in terms of LE and achieves reductions of 8.9% in costs, 25.1% in FP and 20.6% in overdiagnosed cases. In absolute numbers, with an annual discount rate of 3% for every 2,000 women screened, the risk-based strategy Q5074-Q5074-T5074-A5074 would extend about the same number of lives (4) as the uniform B5069 strategy but would avoid 1.5 overdiagnosed cases, 97 FP mammograms (six of them ending with a biopsy) and would save 250,000 euros. The only drawback would be one additional FN. If we consider the uniform strategy B4574 as a reference, the risk-based strategy T5074-T5074-A4574-A4574 results in a 5% higher benefit and reductions of 6.8% in costs, 21.9% in FP and 10.1% in overdiagnosed cases.

Measuring effectiveness with QALYs. Figure 2 and Table S11 in Appendix S1 present the results of the cost-effectiveness and harm-benefit analyses. Table 1 (B) shows that, compared to the B5069 uniform strategy, the risk-based Q5069-Q4574-Q4574-A4074 strategy results in reductions of 8% in costs, 17.2% in FP and 25% in overdiagnosed cases. Similarly, compared to the uniform strategy B4574, the risk-based Q4574-Q4574-A4574-A4074 strategy achieves an increase of 4% in QALYs and reductions of 9.2% in costs, 20.4% in FP and 23% in overdiagnosed cases.

False negative results. We have analyzed the incremental ratios of FN results per unit of benefit separately from the other cost-effectiveness or harm-benefit ratios because the pattern of

changes in FN results is affected differently by the periodicity of the exams and the age-interval of screening. For instance, moving from uniform B5069 to uniform A5069 reduces the amount of FN by 29%, but moving from uniform B5069 to uniform B4574 increases the amount of FN by 33%. Figures 1 and 2 show that there were no strategies in the lower left part of the incremental FN per incremental benefit analyses and the harm-benefit frontier for FN per LE or per QALY only included annual screening strategies.

The last column of Table 1 shows the percentages of changes in FN results for the selected risk-based strategies with respect to the uniform B5069 and B4574 strategies. Compared to the uniform B5069 strategy, the selected risk-based strategies, which have a similar benefit, have more FN results (20% or more when the measure of benefit is LE, 25% or more when the benefit is measured in QALY). Nevertheless, when considering the uniform B4574 strategy, the selected risk-based strategies not only have less FP results and overdiagnosed cases but also have less FN results. The finding that there are more FN results from risk-based screening compared to uniform B5069 than compared to uniform B4574 is mostly due to the fact that, in general, the selected risk-based strategies screen women until age 74.

Summary of optimal strategies. When all the risk-based strategies that are at or near the Pareto frontier are considered and benefit is measured as LE, the risk-based strategies that provide a similar benefit than the B5069 strategy are characterized by quinquennial for the L and ML, triennial for the MH and tri-, bi- or annual periodicities for the H risk groups. When benefit is measured as QALYs, the risk-based strategies are characterized by quinquennial periodicities for the L, ML and MH and annual for the H risk groups. When the standard of comparison is the uniform strategy B4574, the risk-based strategies that provide similar benefits, either LE or QALY, are characterized by quinquennial for the L, triennial for the ML, and annual periodicities for the MH and the H risk groups.

Figures S4 and S5 in section G of Appendix S1 show how the uniform screening strategies, other than B5069 and B4574, performed in the cost-effectiveness and harm-benefit analyses.

Validation of the model inputs

When we assumed a scenario without screening, for the age interval 0 to 74 years, we obtained a cumulative incidence of BC equal to 5.8% and a mortality rate from BC equal to 1.5%. These values were consistent with the literature [36,37]. Section G.1, and Tables S12 and S13 in Appendix S1 compare our results for the biennial strategy B4569 with the results obtained in the INCA study. The detection rates obtained with our model are slightly higher than the INCA rates for both types of detection (screening or interval), except for the 44–49 age group. The overall program sensitivity was very similar (68.1% in the INCA study versus 68.4% in our model). The stage distributions of the models, either screen-detected or interval, were more favorable than the cases in the INCA study. Table 2 shows the distribution of the interval cases in true interval and FN, by time since last mammogram. The timing of overall interval cases and true interval cases was similar in the INCA study and our model. We observed differences in the distribution of FN results at first and second year after the exams. While in the INCA study there was a higher proportion of FN in the second year, our model had a higher proportion of FN in the first year of the interval. Table 3 compares the overall benefit and harm results for the uniform strategies B5069 and B4574 with published reviews [33–35]. We observe similarities between the Lancet review for mortality reduction and with the Cochrane and Euroscreen reviews for overdiagnosis. The ratios of overdiagnosed

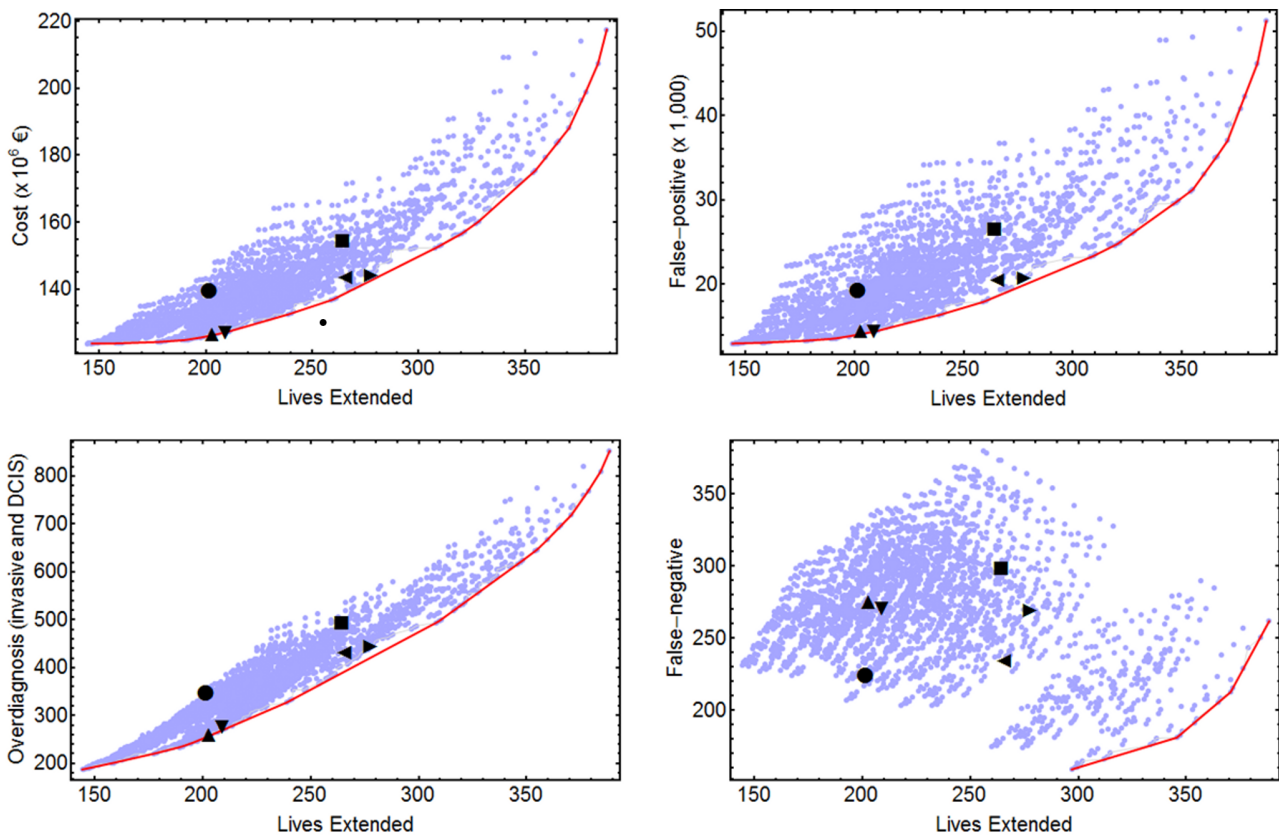


Figure 1. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in lives extended. Dots represent specific screening strategies. Results obtained with an annual discount of 3%. ●: uniform B5069; ■: uniform B4574. ▲: risk-based Q5074-Q5074-Q4574-A4574; ▼: risk-based Q5074-Q5074-T5074-A5074. ◀: risk-based T5069-B5074-A5074-A5074; ▶: risk-based T5074-T5074-A4574-A4574. Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial. The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam. In the risk-based strategies, the four strings correspond to the Low, Medium-Low, Medium-High and High risk groups, respectively.
doi:10.1371/journal.pone.0086858.g001

per LE for the B5069 and B4574 strategies in our study were 1.3 and 1.4, respectively, in the lower range of the reviews.

Sensitivity analysis

Figures S6 and S7 in Appendix S1 show that if there was a migration of women to higher risk groups, the selected risk-based strategies would achieve even higher benefit than the uniform B5069 and B4574 strategies at similar cost and harm values.

Tables S14 and S15 in Appendix S1 present the results of the sensitivity analysis, when the assumptions on the overdiagnosis rates for invasive BC and DCIS, on the costs of cancer treatment, and on the disutility by FP were changed. Tables S14 and S15 also show the relative changes with respect to the uniform B5069 strategy. In general, the cost-benefit and the harm-benefit analyses were robust to changes in the inputs, but we observed changes in the incremental cost-benefit or harm-benefit ratios. When the overdiagnosis rate of invasive or DCIS tumors increased, the incremental cost- or harm-benefit ratios also increased which means that the cost or the harm for each additional unit of benefit was higher. When treatment costs increased, a reduced number of the strategies located in the left part of the frontier were not optimal anymore. This phenomenon was common to both benefit measures (LE and QALY) and was more marked for a 5-fold than for a 2-fold increase. Finally, when the disutility of FP results increased, the optimal strategies were similar, but the incremental

FP per incremental QALY also increased. Section G.2 in Appendix S1 includes further details of the sensitivity analyses.

Discussion

Our analysis aimed to be a global assessment of the impact that a new paradigm of screening would have on benefit, costs and harms rather than a detailed guideline of how personalized screening should be done.

Using probabilistic models, we have found that risk-based screening strategies are more efficient and have lower harm-benefit ratios than uniform strategies. If, instead of screening biennially all women 50 to 69 years old, we combined quinquennial, triennial and annual exam periodicities for women at L or ML, MH, and H risk, respectively, in the age interval 50 to 74, we would avert the same number of deaths. Similarly, strategies that combine quinquennial exams for women at L or ML risk with annual exams for women at MH or H risk, respectively, in the age interval 45 to 74, result in similar gain in QALYs than the uniform biennial strategy in the age interval 45 to 74. But, the important result is that in both cases the risk-based strategies would result in remarkable reductions of costs, FP results and overdiagnosis.

It is important to notice that a risk-based screening strategy Q5074-Q5074-Q4574-A4574 has similar benefits and less costs

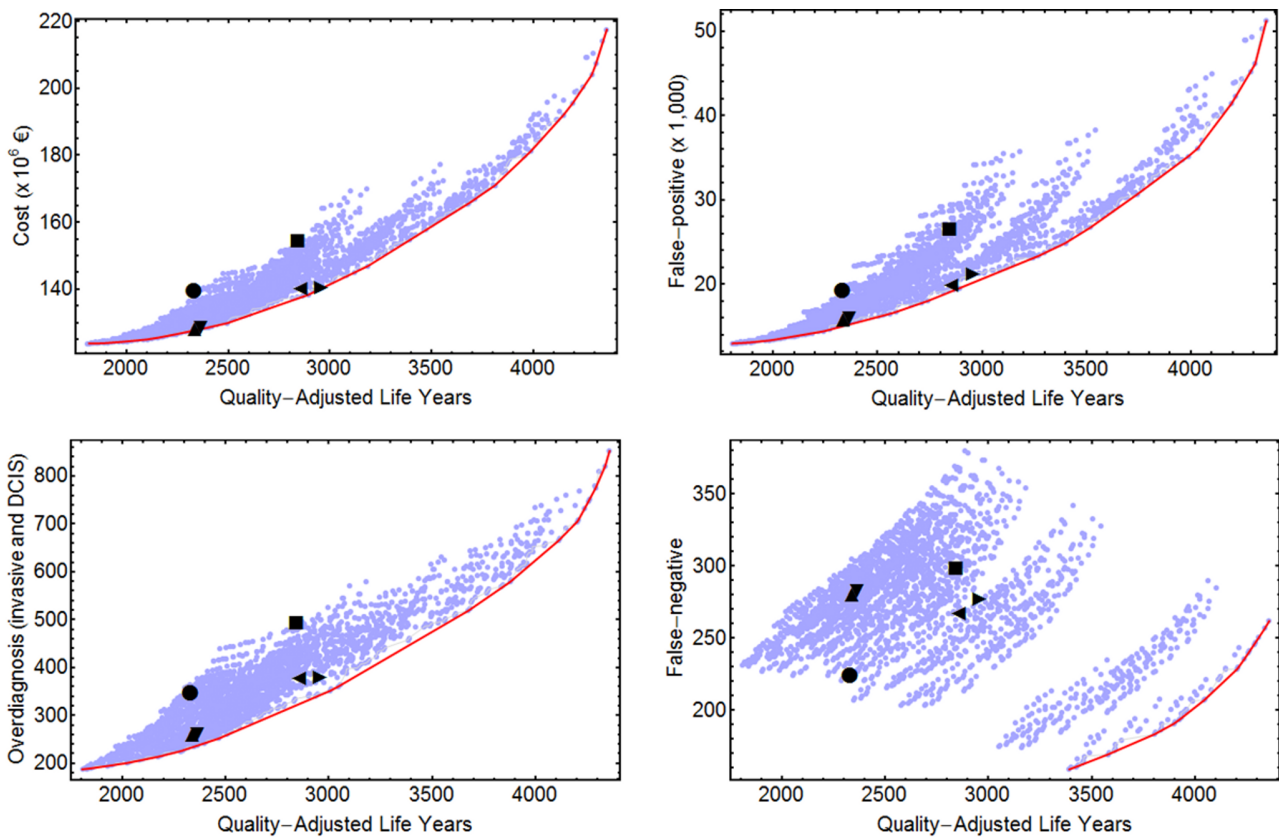


Figure 2. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in quality-adjusted life years. Dots represent specific screening strategies. Results obtained with an annual discount of 3%. ●: uniform B5069; ■: uniform B4574. ▲: risk-based Q5069-Q4574-Q4574-A4574; ▼: risk-based Q5069-Q4574-Q4574-A4074. ◀: risk-based Q5074-Q5074-A4074-A4074; ▶: risk-based Q4574-Q4574-A4574-A4074. Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial. The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam. In the risk-based strategies, the four strings correspond to the Low, Medium-Low, Medium-High and High risk groups, respectively. doi:10.1371/journal.pone.0086858.g002

and harms than the uniform B5069. This does not mean that Q5074-Q5074-Q4574-A4574 should be recommended, only that the same benefits as B5069 can be achieved more efficiently and safely. In fact, in terms of LE, Q5074-Q5074-T5074-A5074 improves the uniform B5069 and has similar costs and harms to Q5074-Q5074-Q4574-A4574. The cost-effectiveness and harm-benefit analyses show the trade-offs when moving along the Pareto frontier. Drawing horizontal lines at the level of uniform strategies, one can estimate the improvement in benefit for a specific cost or harm. Drawing vertical lines allows estimation of the reduction in costs or harms for a specific benefit.

Some recent works have proposed personalized recommendations for BC screening based on cost-effectiveness or cost-utility analyses [8,9] or in decision models that compare harm and benefits [38]. Schousboe *et al.* [8] established cost-effectiveness thresholds of \$100,000 or \$50,000 per QALY gained and compared different periodicities for the screening exams in 10-year age groups, BI-RADS breast density categories and the presence/absence of personal history of biopsy and family history of breast cancer. They recommended that women aged 50 to 79 years who have low breast density and no other breast cancer risk factors may consider having mammography less frequently than every 2 years, which is consistent with our results. But, they recommended biennial screening for women aged 50 to 79 with breast densities of 3 or 4, independently of the presence of the

other two risk factors. In our study, women with breast density 3 or 4 belong to ML, MH, and H risk groups, depending on having 0, 1 or 2 additional risk factors, respectively, and therefore the optimal strategy would have recommended different periodicities and age intervals for these three risk groups. In addition, Schousboe *et al.* concluded that annual mammography was not cost-effective for any group, regardless of age or breast density. These recommendations do not agree with our results, probably due to differences in the studies' objectives and the methodological approaches used.

van Ravesteyn *et al.* used different models - one of them was the LZ model that we used in the present study - to assess the false-positive mammography findings per death averted and per years of life gained in women aged 40 to 49 years [9]. In all models, screening women with increased risk for breast cancer lead to more breast cancer deaths averted with approximately the same number of false-positive results.

Ayer *et al.* [38], using a Markov decision process that considers personal risk characteristics and the personal history of screening, showed that personalized screening strategies outperform the existing guidelines with respect to the total expected quality-adjusted life years, while significantly decreasing the number of mammograms and false-positives. They concluded that screening is less beneficial for most women over age 74 and, as we found,

Table 1. Uniform B5069 and B4574 strategies compared with alternative risk-based strategies.

A) Effect measured in lives extended (LE)¹					
Schedule	LE	Cost (×10 ⁶ €)	False positive ²	Overdiagnosis ³	False negative
Uniform B5069	201.9	139.6	19,256.3	347.6	223.9
Risk-based strategies ⁴	Percentage of change, compared to fixed B5069				
Q5074-Q5074-Q4574-A4574	0.6	-9.3	-25.1	-25.9	22.7
Q5074-Q5074-T5074-A5074	3.8	-8.9	-25.1	-20.6	20.8
Uniform B4574	264.7	154.5	26,578.5	493.1	298.2
Risk-based strategies ⁴	Percentage of change, compared to fixed B4574				
T5069-B5074-A5074-A5074	0.5	-7.7	-23.0	-12.4	-21.6
T5074-T5074-A4574-A4574	5.0	-6.8	-21.9	-10.1	-9.7
B) Effect measured in quality-adjusted life years (QALY)¹					
Schedule	QALY	Cost (×10 ⁶ €)	False positive ²	Overdiagnosis ³	False negative
Uniform B5069	2,333.3	139.6	19,256.3	347.6	223.9
Risk-based strategies ⁴	Percentage of change, compared to fixed B5069				
Q5069-Q4574-Q4574-A4574	0.3	-8.3	-18.3	-25.9	24.9
Q5069-Q4574-Q4574-A4074	1.5	-8.0	-17.2	-25.0	26.2
Uniform B4574	2,848.8	154.5	26,578.5	493.1	298.2
Risk-based strategies ⁴	Percentage of change, compared to fixed B4574				
Q5074-Q5074-A4074-A4074	0.4	-9.2	-25.3	-23.4	-10.5
Q4574-Q4574-A4574-A4074	4.0	-9.2	-20.4	-23.0	-7.2

¹Data correspond to a cohort of 100,000 women at birth assessed in the age-interval 40–79 years. All the absolute values have been discounted at an annual rate of 3%.
²False positive includes both non-invasive and invasive procedures.
³Overdiagnosis of invasive and DCIS cases.
⁴Periodicity and age-interval for Low, Medium-Low, Medium-High and High risk groups, respectively. Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial. The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam.
 doi:10.1371/journal.pone.0086858.t001

provides significant QALY gains, for the high-risk women in the age group 40–49.

Limitations and other considerations

We have used a very detailed model that allowed us to thoroughly assess the cost-effectiveness and harm-benefit of 2,625 different screening scenarios, either risk-based or not. However, our study has several limitations.

First, our model relies on data and assumptions that may be not correct. When available, we have used Catalan or Spanish data from population based registries or BC screening programs. If the input data was not available at the region or country level, we used data that the Cancer Intervention and Surveillance Modeling Network (CISNET) had prepared for BC mortality modeling research groups in the USA, like the distribution of disease stages at diagnosis [12], or from the Breast Cancer Surveillance Consortium, like the distribution of risk factors in the population

Table 2. Distribution of the interval cases by time since last mammogram.

Time since last mammogram (months)	Interval cancer		True interval and minimal signs		False negative and occult tumors	
	N	%	N	%	N	%
The INCA study ¹						
0–11	420	32.4	142	26.2	117	38.7
12–23	876	67.6	399	73.8	185	61.3
Probabilistic model, biennial screening						
0–11	529	35.3	287	26.8	242	56.5
12–23	971	64.7	785	73.2	186	43.5

¹The total number of interval cases in the INCA study is higher than the sum of true interval and FN, occult and minimal signs, because 60.3% of all the interval cases were reviewed.
 doi:10.1371/journal.pone.0086858.t002

Table 3. Comparison with published reviews.

	Our study ¹		Independent UK Panel on Breast Cancer Screening review [34]	Cochrane systematic review ² [33]	Euroscreen review ³ [35]
	B5069	B4574			
Mortality reduction (%)	14.4	19.6	20.0	15.0	23.0–30.0
Deaths averted	4.3	5.8	4.3	0.5	7–9
Overdiagnosis	5.5	8.1	12.9	5.0	4
Non invasive FP	265.5	347.8	-	>100	200
Invasive FP	24.9	28.7	-	-	30
Number needed to screen to extend 1 live	233	172	235	2000	111–143

Benefits and harms per 1,000 women screened.

¹time horizon 40–79 years.²10 years of follow-up.³time horizon 50–79 years.

doi:10.1371/journal.pone.0086858.t003

or the relative risks of the considered risk factors [8,20,22]. Finally, there were some inputs that had been obtained from published randomized clinical trials and observational studies [12,23]. This variety of data sources and modeling assumptions makes it necessary to carefully analyze the model outputs. To validate our model, on one hand, we have performed sensitivity analyses either in this study or in previous publications that show that the model and results are robust to the model assumptions [10,16]. On the other hand, we have reviewed the literature to check whether our results were consistent, at least for the screening strategies that have been included in reviews - mostly biennial strategies in the 50–69 year age interval. The three examined reviews, the Cochrane systematic review [33], the Independent UK Panel on Breast Cancer Screening review [34], and the Euroscreen comprehensive review of European screening programs [35] provide a wide range of values for the benefits and harms of screening. Our results have similarities and differences with the three reviews. We obtained a value close to the Lancet review for number of deaths averted per 1,000 women. Our ratios of overdiagnosed cases per death averted were in the low range of values obtained in the mentioned reviews, 0.5, 3 and 10 overdiagnosed cases per death averted in the Euroscreen, the UK Panel and the Cochrane reviews, respectively. Our estimates of false-positive mammography results were higher than in the reviews, nevertheless for invasive false-positives we were close to the Euroscreen result. Finally, when we compared our results for the uniform screening strategies B4569 or B5069 with the INCA study or other studies of interval cancer [39–41], we found a high consistency in most of the results relative to the number of cancer cases detected per mammography, sensitivity of the program, distribution of screen-detected and interval cases, and distribution of true interval and false-negative cases.

Second, we have assumed that BC risk influenced only the incidence of the disease and not the distribution of stages at diagnosis, the sensitivity and specificity of mammography, the sojourn time in the preclinical state or the mortality from other causes. It could happen that tumors for women at MH or H risk groups had a less favorable stage distribution at diagnosis and the benefit of screening for these groups was lower than estimated. Also, it is known that mammography performance is associated with the considered risk factors [42,43].

Third, we have assumed that there are no changes in the risk factors after the age at which screening exams start. We considered

that the proportion of women in the risk groups remained constant over time and it was the overall sample estimate for the BCSC data. This assumption may not be correct, because as women get older breast density tends to decrease and personal history of biopsy and family history of breast cancer have more chances to be present. We think that our results are robust to changes in the risk group weights over time, as the sensitivity analysis has shown to be the case for changes in the risk group distributions. However, when considering personalized screening, BC risk should be updated when new information on risk factors or their trends is available.

Forth, our model used age-specific sensitivities of the screening exam that correspond to a more prevalent use of film mammography than digital mammography. We did not assess the impact of changing the mammography performance in this study. van Ravesteyn *et al.* [9] found that there was greater harm relative to benefit from digital than from film mammography in women aged 40–49 years, an age group where it seems that digital mammography has higher sensitivity, detects more cases of DCIS and results in more FP results [44,45].

Fifth, our probabilistic model assumes that screening results in a stage-shift at BC diagnosis, but does not consider DCIS as one of the BC stages. Therefore, the fraction of DCIS tumors that would have progressed and been diagnosed as invasive in the absence of screening, are re-distributed under screening in more favorable stages at diagnosis, but not as DCIS. This may have produced an underestimation of the benefit of the screening strategies, both uniform or risk-based. If bias had affected uniform and risk-based strategies similarly, the cost-effectiveness and harm-benefit analyses would remain valid.

We agree with Mandelblatt [46] and Ayer [38] on that risk-based approaches show promise, but there are important issues that need further research. One issue is the need to know more about the underlying relationships between risk factors and the biology of breast cancer and, the other issue, is to overcome the practical issues of implementing appropriate screening strategies based on personalized risk. The PROCAS study in the UK [47], the KARMA project in Sweden [48–50], and the PROSPR network in the USA [51] are examples of advancing towards a tailored screening through improving BC risk prediction. Creating new strategies for communicating individual estimates of benefit and risk of alternative screening methods, to better inform patients and health care providers, is a challenge for researchers.

In conclusion, risk-based screening strategies seem to be more efficient and have better harm-benefit ratios than the standard uniform strategies. We have proposed a reduced number of risk-based screening strategies that combine quinquennial or triennial exams for women in low or moderate-low risk groups and annual exams for women in the moderate-high or high risk groups, for the consideration of researchers, decision makers and policy planners. Now, it is necessary to develop accurate measures of individual risk of BC and to work on how to organise risk-based screening programs.

Supporting Information

Appendix S1 Supporting information on methods used and results obtained, containing Tables S1 to S15 and Figures S1 to S7. Table S1, Distribution of stages at diagnosis of BC. Table S2, Relative risk of breast cancer based on age and breast density. Table S3, Prevalences of risk factors by age group for each category of breast density. Table S4, Characteristics of the 2,625 screening strategies analyzed. Table S5, The utilities for the general population and for women diagnosed with BC, either DCIS or invasive. Table S6, Model for false positives of non-invasive tests. Table S7, Model for false positives of invasive tests. Table S8, Distribution of stages at diagnosis of BC for screen-detected cases. Different overdiagnosis rates. Table S9, Linear regression model with dependent variable being the DCIS rate per 10⁵ women-year. Table S10, Cost-effectiveness and harm-benefit analysis. Lives extended. Table S11, Cost-effectiveness and harm-benefit analysis. Quality-adjusted life years (QALY). Table S12, Number of mammograms and detection rates for screen-detected and interval cases and program sensitivity by age groups. Invasive cancer (DCIS not included). Table S13, Distribution of stages at diagnosis of BC. Table S14, Sensitivity analysis. Changes in lives extended. Table S15, Sensitivity analysis. Changes in QALY. Figure S1, Incidence curves for twelve risk profiles grouped by risk level: (A) Low Risk, (B) Medium-Low Risk, (C) Medium-High Risk, and (D) High Risk. Graphic (E) shows the smoothed incidence rates for each risk group. Figure S2, Observed and smoothed DCIS rates over time in Catalonia (1983–2008). Figure S3, Index of mammography use (IMU) and smoothed DCIS rates over time in Catalonia (1983–2008). Figure S4, Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies, with uniform strategies marked. Effect measured in lives extended. Figure S5, Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies, with uniform strategies marked. Effect measured in QALY. Figure S6, Sensitivity analysis of a change in

the risk groups distribution. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in lives extended. Figure S7, Sensitivity analysis of a change in the risk groups distribution. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in QALY. (PDF)

Acknowledgments

We thank Sandra Lee, Marvin Zelen and Hui Huang for their support and for providing the initial code of the probabilistic models used, Jordi Blanch for data analysis of the INCA study and revising the manuscript and Montserrat Martínez-Alonso for developing the breast cancer incidence and mortality models and revising the manuscript. We are also grateful to JP Glutting for review and editing, to the Cumulative False Positive Research Group (RAFP) project researchers, and to Rebecca Hubbard and two anonymous reviewers for their valuable comments on a preliminary version of the manuscript.

Interval Cancer (INCA) Study Group (alphabetical order): IMIM (Hospital del Mar Medical Research Institute), Barcelona: Jordi Blanch, Xavier Castells, Mercè Comas, Laia Domingo, Francesc Macià, Juan Martínez, Ana Rodríguez-Arana, Marta Román, Anabel Romero, Maria Sala. General Directorate Public Health and Centre for Public Health Research (CSISP), FISABIO, Valencia: Carmen Alberich, Maria Casals, Josefa Ibáñez, Amparo Lluch, Inmaculada Martínez, Josefa Miranda, Javier Morales, Dolores Salas, Ana Torrella. Galician Breast Cancer Screening Program, Xunta de Galicia: Raquel Almazán, Miguel Conde, Montserrat Corujo, Ana Belén Fernández, Joaquín Mosquera, Alicia Sarandeses, Manuel Vázquez, Raquel Zubizarreta. General Directorate of Health Care Programmes. Canary Islands Health Service: Teresa Barata, Isabel Díez de la Lastra, Juana María Reyes. Basque Country Breast Cancer Screening Program. Osakidetza: Arantza Otegi, Garbiñe Sarriugarte. Corporació Sanitària Parc Taulí, Sabadell: Marisa Baré, Núria Torà. Hospital Santa Caterina, Girona: Joana Ferrer, Francesc Castanyer, Gemma Renart. Epidemiology Unit and Girona Cancer Registry; and University of Girona: Rafael Marcos-Gragera, Montserrat Puig-Vives. Universitat de Lleida-IRBLleida: Carles Forné, Montserrat Martínez-Alonso, Albert Roso, Montserrat Rué, Ester Vilapriñó. Universitat Rovira i Virgili, Tarragona: Misericordia Carles, Aleix Gregori, Maria José Pérez, Roger Pla.

Author Contributions

Analyzed the data: EV CF MR. Wrote the paper: MR EV CF. Co-developed the project that includes this study: MR EV MC RP. Developed costs estimations, contributed to the cost-effectiveness analysis: MC. Developed the code for the data analysis: EV CF. Coordinated the RAFP and the INCA projects: MS XC. Participated in the design and analysis of the INCA study: LD. Revised and approved the manuscript: EV CF MC MS RP XC LD MR.

References

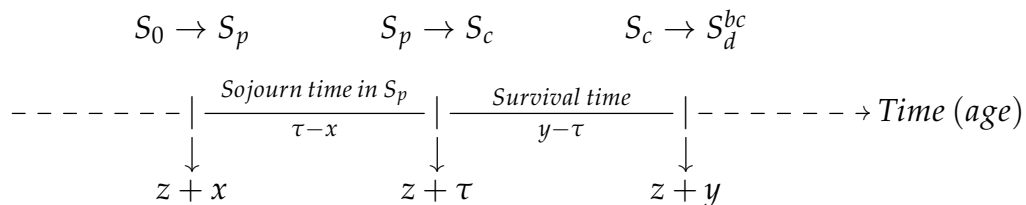
1. Feuer E, Plevritis SK, Berry DA, Cronin KA (eds), et al. (2006) The impact of mammography and adjuvant therapy on US breast cancer mortality (1975–2000): collective results from the Cancer Intervention and Surveillance modeling network. *J Natl Cancer Inst Monogr* (36): 1–126.
2. Roman R, Sala M, Salas D, Ascunce N, Zubizarreta R, et al. (2011) Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol* 23: 104–111.
3. Kalager M, Adami HO, Bretthauer M, Tamimi RM (2012) Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med* 156: 491–499.
4. Jorgensen KJ, Gotzsche PC (2009) Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 339: b2587.
5. Mandelblatt JS, Cronin KA, Bailey S, Berry DA, de Koning HJ, et al. (2009) Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 151: 738–747.
6. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, et al. (2008) European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition-summary document. *Ann Oncol* 19: 614–22.
7. Joy J, Penhoet E, Petitti D, editors, Institute of Medicine and National Research Council Committee on new approaches to early detection and diagnosis of breast cancer (2005) Saving women's lives: strategies for improving breast cancer detection and diagnosis. The National Academies Press, Washington D.C. Available: <http://www.ncbi.nlm.nih.gov/books/NBK22315/>. Accessed 2013 Dec 21.
8. Schousboe JT, Kerlikowske K, Loh A, Cummings SR (2011) Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med* 155: 10–20.
9. van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, et al. (2012) Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. *Ann Intern Med* 156: 609–617.
10. Carles M, Vilapriñó E, Cots F, Gregori A, Pla R, et al. (2011) Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). *BMC Cancer* 11: 192.
11. Lee S, Zelen M (1998) Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *J Am Stat Assoc* (93): 1271–1281.
12. Lee S, Zelen M (2006) A stochastic model for predicting the mortality of breast cancer. *J Natl Cancer Inst Monogr* (36): 79–86.

13. Lee SJ, Zelen M (2008) Mortality modeling of early detection programs. *Biometrics* 64: 386–395.
14. Vilapriño E, Gispert R, Martínez-Alonso M, Carles M, Pla R, et al. (2008) Competing risks to breast cancer mortality in Catalonia. *BMC Cancer* 8: 331.
15. Vilapriño E, Rue M, Marcos-Gragera R, Martínez-Alonso M (2009) Estimation of age- and stage- specific Catalan breast cancer survival functions using US and Catalan survival data. *BMC Cancer* 9: 98.
16. Rue M, Vilapriño E, Lee S, Martínez-Alonso M, Carles M, et al. (2009) Effectiveness of early detection on breast cancer mortality reduction in Catalonia (Spain). *BMC Cancer* 9: 326.
17. Martínez-Alonso M, Vilapriño E, Marcos-Gragera R, Rue M (2010) Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* 12: R58.
18. Zelen M, Feinleib M (1969) On the theory of screening for chronic diseases. *Biometrika* 56: 601–614.
19. Day NE, Walter SD (1984) Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 40: 1–14.
20. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, et al. (2008) Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 148: 337–347.
21. American College of Radiology (2003) The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). American College of Radiology, Reston (VA).
22. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, et al. (2006) Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 98: 1204–1214.
23. Lidgren M, Wilking N, Jonsson B, Rehnberg C (2007) Health related quality of life in different states of breast cancer. *Qual Life Res* 16: 1073–1081.
24. Biesheuvel C, Barratt A, Howard K, Houssami N, Irvig L (2007) Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol* 8: 1129–1138.
25. de Koning HJ, Draisma G, Fracheboud J, de Brujin A (2006) Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. *Breast Cancer Res* 8: 202.
26. Morrell S, Barratt A, Irvig L, Howard K, Biesheuvel C, et al. (2009) Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control* (21): 275–282.
27. Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP (2006) Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 332: 689–692.
28. Zahl PH, Maehlen J, Welch HG (2008) The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 168: 2311–6.
29. Rue M, Carles M, Vilapriño E, Martínez-Alonso M, Espinas JA, et al. (2008) Dissemination of periodic mammography and patterns of use, by birth cohort, in Catalonia (Spain). *BMC Cancer* 8: 336.
30. Perez MJ, Gregori A, Carles M, Gispert R, Martínez-Alonso M, et al. (2010) The evolution of breast cancer mortality and the dissemination of mammography in Catalonia: an analysis by health region. *Rev Esp Salud Publica* 84: 691–703.
31. Erbas B, Amos A, Fletcher A, Kavanagh AM, Gertig DM (2004) Incidence of invasive breast cancer and ductal carcinoma in situ in a screening program by age: should older women continue screening? *Cancer Epidemiol Biomarkers Prev* 13: 1569–1573.
32. Lopez Bastida J, Oliva J, Antonanzas F, Garcia-Altes A, Gispert R, et al. (2010) A proposed guideline for economic evaluation of health technologies. *Gac Sanit* 24: 154–170.
33. Gotzsche PC, Jorgensen K (2013) Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 6: CD001877.
34. Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. *Lancet* 380: 1778–1786.
35. Paci E, Euroscreen Working Group (2012) Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 19 Suppl 1: 5–13.
36. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, et al. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 127: 2893–2917.
37. Izquierdo A, Gispert R, Saladie F, Espinas JA (2008) Analysis of cancer incidence, survival and mortality according to the main tumoral localizations, 1985–2019: Breast cancer. *Med Clin (Barc)* 131 Suppl 1: 50–52.
38. Ayer T, Alagoz O, Stout N (2012) A POMDP Approach to Personalize Mammography Screening Decisions. *Operations Research* 60: 1019–1034.
39. McCann J, Britton PD, Warren RM, Hunnam G (2001) Radiological peer review of interval cancers in the East Anglian breast screening programme: what are we missing? East Anglian Breast Screening Programme. *J Med Screen* 8: 77–85.
40. Domingo L, Sala M, Servitja S, Corominas JM, Ferrer F, et al. (2010) Phenotypic characterization and risk factors for the main tumoral localizations, in a population-based breast cancer screening program in Barcelona, Spain. *Cancer Causes Control* 21: 1155–1164.
41. Kirsh VA, Chiarelli AM, Edwards SA, O'Malley FP, Shumak RS, et al. (2011) Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program. *J Natl Cancer Inst* 103: 942–950.
42. Kerlikowske K, Carney PA, Geller B, Mandelson MT, Taplin SH, et al. (2000) Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Ann Intern Med* 133: 855–863.
43. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, et al. (2003) Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 138: 168–175.
44. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, et al. (2005) Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 353: 1773–1783.
45. Karssemeijer N, Bluekens AM, Beijerinck D, Deurenberg JJ, Beekman M, et al. (2009) Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 253: 353–358.
46. Mandelblatt JS, Stout N, Trentham-Dietz A (2011) To screen or not to screen women in their 40s for breast cancer: is personalized risk-based screening the answer? *Ann Intern Med* 155: 58–60.
47. National Health Service (no date) The PROCAS Study website. Available: <http://www.uhsm.nhs.uk/research/Pages/PROCASstudy.aspx>. Accessed 2013 Dec 21.
48. Karolinska Institutet (no date) Karolinska Mammography Project for Risk Prediction of Breast Cancer website. Available: <http://karmastudy.org/>. Accessed 2013 Dec 21.
49. Darabi H, Czene K, Zhao W, Liu J, Hall P, et al. (2012) Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res* 14: R25.
50. Li J, Szekeley L, Eriksson L, Hedddson B, Sundbom A, et al. (2012) High-throughput mammographic density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Res* 14: R114.
51. National Cancer Institute (no date) Cancer Control Research website. Available: <http://cancercontrol.cancer.gov/grants/abstract.asp?applied=8223497>. Accessed 2013 Dec 21.

Appendix S1

A The probabilistic model

Lee and Zelen (LZ) developed a probabilistic model that predicts mortality as a function of the early detection modality. The characteristics and assumptions of the LZ model are described in detail elsewhere [1-3]. The assumptions of the LZ model are (1) a four-state progressive disease in which a subject may be in a disease-free state (S_0), preclinical disease state (S_p : capable of being diagnosed by a special exam), clinical state (S_c : diagnosis by symptomatic detection), and a death from breast cancer (BC) state (S_d); (2) age-dependent transitions into the different states; (3) age-dependent examination sensitivity; (4) age-dependent sojourn times in each state; and (5) exam-diagnosed cases have a stage-shift in the direction of more favorable prognosis relative to the distribution of stages in symptomatic detection.



The LZ model considers:

- n screening exams at times $t_0 < t_1 < \dots < t_{n-1}$. It is assumed that $t_0 = 0$ and $z = \text{age}$ at t_0 .
- Three chronological times (see schema):
 - x : time at entering S_p , $z + x$: age when entering S_p . The time x is not observed but can be derived from the incidence function and the distribution of sojourn time in the S_p state. x takes a negative value if the transition to S_p occurs before the age at first exam, z .
 - τ : time at entering S_c , $z + \tau$: age at entering S_c . The time τ can not be observed in cases detected by exam, only in the clinically detected cases. For cases detected by exam, τ can be estimated.
 - y : time at death, $z + y$: age at death. Then $x < \tau < y$
- Sojourn time in S_p : $\tau - x$
- Sojourn time in S_c : $y - \tau$

The LZ basic model calculates the cumulative probability of death for the cohort group exposed to any screening program after T years of follow-up. Similarly, the cumulative probability of death for the cohort group receiving typical health care can be calculated. These probabilities are used to calculate

possible reductions in mortality in an early detection program after T years of follow-up.

Survival distributions for exam-diagnosed, interval, and control cases are assumed to be conditional on the stage at diagnosis and treatment, but are not dependent on the mode of diagnosis. The LZ model assumes k stages, $\phi_s(j)$, $\phi_i(j)$ and $\phi_c(j)$ represent the probability of being diagnosed at stage j , $j = 1, \dots, k$ for exam-diagnosed, interval and control cases, respectively, and $f_j(t|z + \tau)$ is the probability density function (pdf) of survival time t among subjects who would have been clinically diagnosed at stage j in the absence of screening. Then the survival time *pdfs* of the exam-diagnosed, interval and control cases are the mixtures: $g_s(t|z + \tau) = \sum_{j=1}^k \phi_s(j) f_j(t|z + \tau)$, $g_i(t|z + \tau) = \sum_{j=1}^k \phi_i(j) f_j(t|z + \tau)$ and $g_c(t|z + \tau) = \sum_{j=1}^k \phi_c(j) f_j(t|z + \tau)$, respectively.

Since screening will appear to increase survival time, the LZ model controls for *lead time* bias by setting the origin of survival time for the screened, interval, and clinical cases at the time of clinical diagnosis. Consequently, there is an implied *guarantee time* for disease-specific survival, that is, the cases diagnosed earlier would have been alive at the time the disease would have been clinically diagnosed. This guarantee time, also called *lead time*, is a random variable and is incorporated into the equations of the model. Explicitly, the lead time is $\tau - t_r$ where τ is the time at which the individual enters the clinical state and t_r is the time at which the r detection exam, when the disease will be diagnosed, is given.

A.1 Survival functions

Since the BC survival functions in presence of screening are affected by the lead time and length biases, we used the Catalan BC survival functions by age and stage for the period 1980-89 [4], which correspond to the pre-screening era in Spain. To introduce the benefit of BC treatments during the 1990s and the 2000s, we multiplied the 1980-89 BC hazard rates by stage-specific hazard ratios (HR) found in the literature. For local and regional BC, the hazard ratios of adjuvant treatment—either multi-chemotherapy or hormonal treatment—were adapted from Mariotto *et al.* [5]. We assumed HR=0.85 for localized BC (AJCC stages I or II-) and HR=0.8 for regional BC (stages II+ and III). For metastatic BC we assumed HR=0.72 [6].

A.2 Distribution of stages at diagnosis

Table S1 presents the distribution of stages at diagnosis for screen-detected, interval and background cancer cases, according to periodicity of exams. The stage distribution for cases without screening was obtained from the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute and the stage distribution for the screen-detected and interval cases from the Breast Cancer Surveillance Consortium (BCSC), by Lee and Zelen [2]. For screen-detected cases and interval cases Lee and Zelen distinguished between annual, biennial and irregular screening.

The stage distributions of screen-detected cases in Table S1 were modified to deal with assumptions of overdiagnosis in the main analysis as well as in the sensitivity analysis. (See section E and Table S8 of this Appendix S1).

B Invasive BC incidence by risk group

B.1 Overall incidence model for invasive BC

To estimate the age-specific incidence rates for invasive BC we used the model published elsewhere by Martinez-Alonso *et al.* [7]. This model estimates the expected number of incident cases, $E(I)$, using the equation:

$$E(I) = \exp(-6.0626 - 38.4818 \text{ Age}_1 + 0.0005 \text{ Age}_2 + 0.6250 \text{ PM50}_1 + 0.0120 \text{ YB}_1 + \text{offset}),$$

where $\text{Age}_1 = (\text{age}/10)^{-2} - 0.0331$, $\text{Age}_2 = (\text{age}/10)^3 - 166.375$, $\text{YB}_1 = \text{year-of-birth} - 1937.5$, $\text{PM50}_1 = \text{PM50}^{0.5} - 0.4342$ (with PM50 being the proportion of women who were having periodic mammograms for early detection at age 50) and $\text{offset} = \log(\text{exposed}) - 0.15 \text{ CFR}$, where CFR indicates the complete fertility rate of the cohort. To obtain the background incidence we assumed $\text{PM50}=0$.

B.2 BC incidence in risk groups

BC risk groups were defined according to the following variables: breast density (measured using the BI-RADS categories 1 to 4 [8]), family history of BC in first degree relatives (yes/no) and personal history of breast biopsy (yes/no). Relative risks of invasive breast cancer based on breast density, family history of BC, and breast biopsy were taken from Tice *et al.* [9] and Schousboe *et al.* [10].

To obtain the age-specific BC incidence by risk group we took the following steps:

1. We estimated the age-specific BC incidence for each breast density category based on the reported relative risks from Schousboe *et al.* [10]. Table S2 shows, by age groups, the relative risks for each breast density level compared to all women in the corresponding age group.
2. Based on data from Tice *et al.* [9], we approximated the relative risks of invasive breast cancer to 1.5 or 0.9, for the presence or absence of a family history of BC in a first degree relative, respectively. The same relative risks, 1.5 or 0.9, were assumed for a positive or negative history of prior breast biopsy.
3. With the previous considerations, twelve detailed age-specific incidence curves were obtained. We grouped them in four aggregated age-specific incidence curves taking into account the proximity of the detailed curves. The four incidence groups represented: (i) category 1

breast density with at most one risk factor, and category 2 breast density with no risk factors (Low Risk); (ii) category 1 breast density with two risk factors, and category 2 breast density with one risk factor, and category 3, 4 breast density with no risk factors (Medium-Low Risk); (iii) category 2 breast density with two risk factors, and category 3, 4 breast density with one risk factor (Medium-High Risk); and category 3, 4 breast density with two risk factors (High Risk).

4. The incidence rates of the four aggregated risk groups were estimated as weighted sums of the detailed incidence curves. The weights were based on the prevalences of each combination of risk factors obtained from the Risk Estimation Dataset of the Breast Cancer Surveillance Consortium (BCSC) [11]. Figure S1 shows the detailed and aggregated incidence curves.

Table S3 shows the prevalences of risk factors by age group for each level of breast density.

C Quality Adjusted Life Years (QALY)

The QALYs for the different breast cancer disease states were obtained using the health related quality of life (HRQoL) measures from Lidgren *et al.* [12]. As Schousboe *et al.* [10], we used the HRQoL measures obtained from the EuroQoL EQ-5D self classifier in the Lidgren *et al.*'s study, which provided HRQoL measures for the first year after primary BC (EQ-5D=0.696); the first year after recurrence (EQ-5D=0.779); the second and following years after primary BC or recurrence (EQ-5D=0.779); and the metastatic BC state (EQ-5D=0.685).

We also followed Schousboe *et al.* [10] to obtain the QALY values for the first year after BC diagnosis: 0.744 for ductal carcinoma in situ (DCIS), 0.696 for localized BC and 0.620 for regional or distant BC. For women that did not die of BC, beyond the first year of diagnosis and during four years, we assumed no loss in quality of life for women with DCIS and a QALY value of 0.779 for the invasive cancers, either local, regional or distant at diagnosis. For women that died of breast cancer, we considered that the last four years of their lives were spent in a distant stage with a QALY value of 0.685, independently of the stage at diagnosis.

The HRQoL data of the Lidgren *et al.*'s study were not reported by age. To obtain the QALY values by age group and BC state we replicated the work of Schousboe *et al.* [12]. The estimated average QALY value of Swedish women at the mean age of the Lidgren's population was 0.823. The proportion of expected quality of life, for any given age for all of the BC states, compared to perfect health (QALY = 1.0) was obtained dividing the previously stated values for each BC state by 0.823. The final QALY value, for each BC state and for any given age group, was obtained multiplying these proportions by the QALY value for that age group in the general population. The QALY values that we used are shown in Table S5.

For false-positive mammograms we assumed an average annualized loss of quality of life of 0.013. Schousboe *et al.* [10] reported that they used this value in their sensitivity analysis but assumed no loss in quality of life in their main analysis. To obtain the 0.013 value they assumed that 50% of women with a false-positive would have anxiety sufficient to increase the mood subscale of the EuroQol instrument from 0 to 1, lasting a total of 2 months. According to the U.S. EQ-5D tariffs, such a change for an entire year represents a decrease in the QALY value of 0.156, which divided by 12 (1/2 women \times 1/6 years) is 0.013. In the sensitivity analysis we assessed the impact of changing the disutility by false-positive to 0, and to 0.026.

D Model that estimates the false-positive rates for invasive and non-invasive tests

The Cumulative False Positive Research study (RAFP) provided the false-positive (FP) rates for invasive and non-invasive tests for women starting biennial exams at ages 44-45 to 68-69. The maximum number of screening rounds was six (non-published data).

We used the FP rate as the dependent variable and included the screening round and age as independent variables, in a log linear model. Tables S6 and S7 contain the parameter estimates. The model allowed us to estimate the FP rate for annual screening. We assumed that after the 6th exam the FP rate was constant. Tables S6 and S7 show the estimated coefficients, confidence intervals and p-values.

E Distribution of stages for screen-detected cases under the assumption of overdiagnosis of invasive BC

To re-calculate the stages' distribution of screen-detected cases, first we subtracted the overdiagnosis rate (e.g 15% for the main analysis) from the proportion of stage I of BC. Then, we divided the proportions of each stage by 1–overdiagnosis rate and obtained the new distribution of stages' distribution. Table S8 presents the stages' distributions for screen detected cases used in the main analysis as well in the sensitivity analyses.

As an example, the stages' distribution for annual screen detected cases under the assumption of 15% overdiagnosis rate, for the age-group 40-49 years, has been obtained as follows:

- From Table S1 the proportion in stage I is 0.62. If we subtract 0.15, we obtain a proportion of $0.62 - 0.15 = 0.47$ women in stage I.
- We subtract 0.15 from 1 and obtain a total of 0.85 women diagnosed with invasive cancer.
- We divide the new proportion for stage I and the old proportions for stages II to IV by 0.85. We obtain $0.47/0.85 = 0.5529$ women in stage I, $0.1131/0.85 = 0.1331$ in stage II–, $0.2141/0.85 = 0.2519$ in stage II+,

$0.0436/0.85 = 0.0513$ in stage III and $0.0092/0.85 = 0.0108$ in stage IV, as indicated in Table S8.

F DCIS attributable to screening

Figure S2 shows the age-specific DCIS rates over time, using the data from the Cancer Registries of the Girona and Tarragona provinces in Catalonia. We used natural splines to obtain a smooth trajectory of the DCIS rates in the age groups 40-49, 50-69 and 70 or more years.

An increase in the DCIS rates, starting at the beginning of the 1990s was observed in the three age groups. In the 40-49 and 50-69 year age groups, we observed a plateau near the year 2000 followed by a new increase afterwards. The DCIS in the older age group seemed to have stabilized during the 2000s.

Our purpose was to quantify the increase in the DCIS rate in relation to the dissemination of mammography for early detection. First, we estimated an index of mammography use (IMU). The data was obtained, for the Girona and Tarragona provinces, from three health surveys performed in the years 1994, 2002 and 2006. Data from these surveys was already analyzed to study the dissemination of periodic mammography in Catalonia [13, 14]. For each health survey year, the IMU was obtained using the following formula:

$$IMU = \sum_{i \in \{1,2,4\}} f_i \times 1/i,$$

where f_i is the fraction of women in the population that received early detection mammograms with a periodicity of i years, with $i = 1, 2, 4$ if the interval between mammograms was 1, 2 or 3 or more years, respectively. The IMU measures the intensity of mammography use per woman-year. If no women in the population used periodic mammography, $IMU=0$. If every woman in the population received an annual mammography, $IMU=1$ and if half of the women in the population receive a biennial mammograph, $IMU= 0.25$.

Then, we used a linear regression model to assess the impact of mammography use in the DCIS rate. Age-group was included in the model as a categorical variable. For this model, we used the DCIS rates corresponding to a baseline period (years 1983-1987) and three periods of three years centered around the health survey years, 1993-1995, 2001-2003 and 2005-2007. For the baseline period, when there was very little use of mammography, it was assumed that the $IMU=0$.

Figure S3 presents the smoothed DCIS rates and the IMU values. It can be observed that the pattern of increase of the DCIS rate is related to the pattern of growth of the IMU, mainly for the 50-69 age group, which is the target population for the public screening program in Catalonia.

Table S9 presents the results of the linear regression model. The IMU is associated with the DCIS rate with a beta coefficient equal to 31.13, indicating an increase of approximately 31.13 DCIS cases per 100,000 women-year if the IMU changes from 0 (no screening) to 1 (all women receive a mammography). The R^2 coefficient of determination is 0.87, denoting that the model explains a high percentage of the total variability in the DCIS rate.

G Results

Tables S10 and S11 show the Pareto efficiency frontiers of cost-effectiveness and harm-benefit analyses in terms of lives extended (LE) and quality-adjusted life years gained (QALY), respectively.

Figures S4 and S5 contain the same data as Figures 1 and 2 with all the studied uniform strategies labeled.

G.1 Validation of the model inputs

Table S12 compares the screen-detected and interval cancer rates per 1,000 mammograms in the INCA study and the fixed B4569 strategy of the mathematical model. The model detection rates are slightly higher than the INCA rates for both types of detection (screening or interval), except for the 44-49 age group. The age-specific interval rates for the INCA study do not show an increasing pattern as the model interval rates.

The overall program sensitivity was very similar (68.1% in the INCA study versus 68.4% in the mathematical model). By age group, both the INCA study and the mathematical model show an increasing trend (Table S12).

Table S13 presents the stages distribution for screen-detected and interval cases. Screen-detected cases in the model had a higher distribution of cases in Stages I and II than the INCA study, which had a higher distribution of Stage III cases. Similarly, interval cases in the INCA study had a higher proportion of Stages III and IV cases. In summary, the stage distributions of the model, either screen-detected or interval, were more favorable than the cases in the INCA study.

G.2 Sensitivity analysis

G.2.1 Changes in the risk groups distribution

Figures S6 and S7 present the cost-effectiveness and harm-benefit analyses when 20% of the women in the L, ML, and MH risk groups migrated to the next higher risk group, for the selected risk-based strategies compared to the uniform B5069 and B4574 ones. With the new risk groups distribution, the selected risk-based strategies even perform better than with the former risk groups distribution.

G.2.2 *Changes in the overdiagnosis rate of invasive tumors*

Tables S14 and S15 present the results obtained, for the benefit measures LE and QALY, respectively, when the assumptions on overdiagnosis rates for invasive BC and DCIS and the cost of cancer treatment were changed. In Table S14 the absolute values correspond to the selected Q5074-Q5074-T5074-A5074 strategy whereas in Table S15 they correspond to the selected Q5069-Q4574-Q4574-A4074 strategy. In both tables the percentages in parentheses correspond to relative changes with respect to the fixed B5069 strategy.

In the main analysis an overdiagnosis rate of 15% was assumed for the invasive tumors when the mammography exams were performed. The sensitivity analysis showed that, in absolute terms, the measures of benefit (LE and QALY) decreased when the overdiagnosis rate increased from 0% to 5% and from 5% to 15%. When the overdiagnosis rate increased from 15% to 25% the measures of benefit (LE and QALY) increased for the less intensive strategies and decreased for the more intensive strategies. The cost-benefit and the harm-benefit analyses were robust to changes in the overdiagnosis of invasive tumor rates. The resultant optimal strategies were mostly the same as in the main analysis but we observed changes in the incremental cost-benefit or harm-benefit ratios. Changes were more marked when the overdiagnosis rate increased from 15% to 25% where some less intensive screening strategies exited the Pareto frontier. In general, when the overdiagnosis rate increased, the incremental cost- or harm-benefit ratios also increased which means that the cost or the harm for each additional unit of benefit was higher.

For the risk-based Q5074-Q5074-T5074-A5074 strategy, increasing the overdiagnosis rate of invasive tumors led to a decrease in the number of LE (Table S14), but the percentage change with respect to the fixed B5069 strategy increased. Similarly, for the risk-based Q5069-Q4574-Q4574-A4074 strategy, increasing the overdiagnosis rate of invasive tumors led to a decrease in the number of QALYs (Table S15) but, compared to the fixed B5069 strategy, the percentage change also increased.

G.2.3 *Changes in the overdiagnosis rate of DCIS*

The overdiagnosis rate of DCIS is directly associated with a slight increase in costs (Tables S14 and S15) and a slight decrease in QALYs (Table S15). Since we assumed that overdiagnosed DCIS did not die of breast cancer, both LE and QALY were not affected by changes in the overdiagnosis rate. The resultant optimal strategies were the same as for the main analysis, except for one of the optimal strategies in the FN-QALY harm-benefit analysis.

For the risk-based Q5074-Q5074-T5074-A5074 and Q5069-Q4574-Q4574-A4074 strategies, increasing the overdiagnosis rate of DCIS led to an increase in costs and in the percentage change with respect to the fixed B5069 strategy (Tables S14 and S15).

G.2.4 Changes in the costs of breast cancer treatment

When treatment costs increase, a reduced number of the strategies located in the left part of the frontier were not optimal anymore. This phenomenon was common to both benefit measures (LE and QALY) and was more marked for a 5-fold than for a 2-fold increase.

G.2.5 Changes in the disutility of FP results

This sensitivity analysis only has an impact on QALYs. Increasing the disutility of FP results led to a decrease in QALY (Table S15).

Tables

Table S1: Distribution of stages at diagnosis of breast cancer

Age (years)	Stages ¹				
	I	II-	II+	III	IV
Background ^{1,2}					
40-49	0.3008	0.2277	0.3091	0.0999	0.0625
50-59	0.2868	0.2176	0.3111	0.1021	0.0825
60-69	0.3028	0.2225	0.2713	0.0974	0.1061
70-79	0.3157	0.2671	0.2227	0.0983	0.0961
Annual screening. Screen-detected cases ^{1,3}					
40-49	0.6200	0.1131	0.2141	0.0436	0.0092
50-59	0.6669	0.1057	0.1935	0.0296	0.0043
60-69	0.7641	0.0739	0.1412	0.016	0.0047
70-79	0.7821	0.0875	0.1067	0.0165	0.0072
Annual screening. Interval cases ^{1,3}					
40-49	0.4644	0.1903	0.2598	0.0667	0.0188
50-59	0.4501	0.1744	0.2976	0.0665	0.0113
60-69	0.5417	0.1532	0.2320	0.0591	0.0141
70-79	0.5446	0.2345	0.1583	0.0496	0.013
Biennial screening. Screen-detected cases ^{1,3}					
40-49	0.5839	0.1217	0.2360	0.0438	0.0146
50-59	0.6210	0.1472	0.1734	0.0423	0.0161
60-69	0.6563	0.1295	0.1830	0.0246	0.0067
70-79	0.7287	0.1311	0.1128	0.0137	0.0137
Biennial screening. Interval cases ^{1,3}					
40-49	0.3673	0.2246	0.3099	0.0819	0.0164
50-59	0.2945	0.2609	0.2648	0.1166	0.0632
60-69	0.4077	0.2231	0.2672	0.0744	0.0275
70-79	0.4336	0.2885	0.1770	0.0673	0.0336
Triennial or quinquennial screening. Screen-detected cases ^{1,3}					
40-49	0.4910	0.1576	0.2605	0.0614	0.0295
50-59	0.5096	0.1707	0.2193	0.0622	0.0382
60-69	0.5384	0.1605	0.2124	0.0489	0.0398
70-79	0.5910	0.1765	0.1494	0.0419	0.0412
Triennial or quinquennial screening. Interval cases ^{1,3}					
40-49	0.3466	0.2262	0.3097	0.0868	0.0307
50-59	0.2919	0.2465	0.2802	0.1118	0.0696
60-69	0.3727	0.2229	0.2686	0.0821	0.0537
70-79	0.3943	0.2814	0.1922	0.0776	0.0545

¹ American Joint Committee on Cancer (AJCC) stage distribution.

² From Surveillance, Epidemiology, and End Results (SEER).

³ From Breast Cancer Surveillance Consortium (BCSC).

Table S2: Relative risk of breast cancer based on age and breast density (the reference group for each relative risk is all women of that age group)¹

BI-RADS category	Age groups				
	40-49	50-59	60-64	65-69	70-79
1	0.351	0.388	0.400	0.581	0.600
2	0.730	0.807	0.832	0.885	0.914
3	1.131	1.251	1.291	1.228	1.268
4	1.468	1.623	1.675	1.283	1.325

¹ From Schousboe *et al.* [10].

BI-RADS (American College of Radiology Breast Imaging Reporting and Data System) 1 = almost entirely fat; 2 = scattered fibroglandular densities; 3 = heterogeneously dense; 4 = extremely dense.

Table S3: Prevalences of risk factors by age group for each category of breast density¹

	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
BI-RADS 1										
0 RF	0.0251	0.0343	0.0403	0.0521	0.0644	0.0778	0.0859	0.0920	0.0932	0.0943
1 RF	0.0128	0.0078	0.0101	0.0143	0.0198	0.0241	0.0294	0.0330	0.0341	0.0346
2 RF	0.0011	0.0006	0.0009	0.0015	0.0023	0.0030	0.0040	0.0036	0.0045	0.0042
BI-RADS 2										
0 RF	0.1931	0.2604	0.2746	0.3089	0.3336	0.3461	0.3477	0.3490	0.3512	0.3501
1 RF	0.1182	0.0699	0.0825	0.1016	0.1234	0.1386	0.1529	0.1626	0.1594	0.1581
2 RF	0.0136	0.0062	0.0083	0.0120	0.0164	0.0198	0.0211	0.0233	0.0248	0.0246
BI-RADS 3										
0 RF	0.2589	0.3452	0.3226	0.2850	0.2448	0.2149	0.1946	0.1812	0.1792	0.1806
1 RF	0.1826	0.1109	0.1191	0.1216	0.1201	0.1144	0.1098	0.1051	0.1015	0.0974
2 RF	0.0225	0.0111	0.0150	0.0168	0.0178	0.0170	0.0173	0.0168	0.0180	0.0173
BI-RADS 4										
0 RF	0.0904	0.1060	0.0826	0.0537	0.0341	0.0258	0.0220	0.0194	0.0200	0.0241
1 RF	0.0719	0.0424	0.0384	0.0285	0.0202	0.0160	0.0132	0.0117	0.0118	0.0125
2 RF	0.0097	0.0051	0.0056	0.0039	0.0033	0.0023	0.0021	0.0022	0.0024	0.0023

¹ Based on the Risk Estimation Dataset of the Breast Cancer Surveillance Consortium (BCSC) [11].

BI-RADS (American College of Radiology Breast Imaging Reporting and Data System) 1 = almost entirely fat; 2 = scattered fibroglandular densities; 3 = heterogeneously dense; 4 = extremely dense.

RF = Risk Factor(s).

Table S4: Characteristics of the 2,625 screening strategies analyzed

Strategy	Interval between exams according to risk group (years)				Age at starting according to risk group (years)				Age at ending according to risk group (years)			
	L	M-L	M-H	H	L	M-L	M-H	H	L	M-L	M-H	H
1	5	5	5	5	50	50	50	50	69	69	69	69
2	5	5	5	5	50	50	50	50	69	69	69	74
3	5	5	5	5	50	50	50	50	69	69	74	74
4	5	5	5	5	50	50	50	50	69	74	74	74
5	5	5	5	5	50	50	50	50	74	74	74	74
6	5	5	5	5	50	50	50	45	74	74	74	74
...
...	5	3	2	1	50	50	45	40	69	69	74	74
...
2625	1	1	1	1	40	40	40	40	74	74	74	74

L = Low risk; M-L = Medium-Low risk; M-H = Medium-High risk; H = High risk.

Table S5: The utilities for the general population and for women diagnosed with breast cancer, either DCIS or invasive¹

Age group	Women in the general population	DCIS first year	Local first year	Regional or distant first year	L, R or D successive years ²	Four years before death
40-49	0.859	0.777	0.726	0.647	0.813	0.715
50-59	0.845	0.764	0.715	0.637	0.800	0.703
60-69	0.812	0.734	0.687	0.612	0.769	0.676
70-79	0.788	0.712	0.666	0.594	0.746	0.656

¹ Adapted from Lidgren *et al.* [12] and Schousboe *et al.* [10].

² QALY for DCIS in successive years are the same as for the general population.

L = local, R = regional, D = distant.

Table S6: Model for false positive results of non-invasive tests

Variables	Coefficient	SE	p-value
Intercept	-1.2905	0.1857	<0.001
Screening round ¹	-0.1424	0.0160	<0.001
Age at screening	-0.0322	0.0031	<0.001
First screening round (<i>yes</i> versus <i>no</i>)	0.5520	0.0653	<0.001

¹ Screening round ranges from 1 to 6.

Table S7: Model for false positive results of invasive tests

Variables	Coefficient	SE	p-value
Intercept	-4.6539	0.1878	<0.001
Screening round ¹	-0.1483	0.0162	<0.001
Age at screening	-0.0195	0.0032	<0.001
First screening round (<i>yes</i> versus <i>no</i>)	1.0638	0.0660	<0.001

¹ Screening round ranges from 1 to 6.

Table S8: Distribution of stages at diagnosis of breast cancer for screen-detected cases. Different overdiagnosis rates¹

Age (years)	Stages				
	I	II-	II+	III	IV
Annual screening					
Overdiagnosis rate 5%					
40-49	0.6000	0.1191	0.2254	0.0459	0.0097
50-59	0.6494	0.1113	0.2037	0.0312	0.0045
60-69	0.7517	0.0778	0.1486	0.0168	0.0049
70-79	0.7706	0.0921	0.1123	0.0174	0.0076
Overdiagnosis rate 15%					
40-49	0.5529	0.1331	0.2519	0.0513	0.0108
50-59	0.6081	0.1244	0.2276	0.0348	0.0051
60-69	0.7225	0.0869	0.1661	0.0188	0.0055
70-79	0.7436	0.1029	0.1255	0.0194	0.0085
Overdiagnosis rate 25%					
40-49	0.4933	0.1508	0.2855	0.0581	0.0123
50-59	0.5559	0.1409	0.2580	0.0395	0.0057
60-69	0.6855	0.0985	0.1883	0.0213	0.0063
70-79	0.7095	0.1167	0.1423	0.0220	0.0096
Biennial screening					
Overdiagnosis rate 5%					
40-49	0.5620	0.1281	0.2484	0.0461	0.0154
50-59	0.6011	0.1549	0.1825	0.0445	0.0169
60-69	0.6382	0.1363	0.1926	0.0259	0.0071
70-79	0.7144	0.1380	0.1187	0.0144	0.0144
Overdiagnosis rate 15%					
40-49	0.5105	0.1432	0.2776	0.0515	0.0172
50-59	0.5541	0.1732	0.2040	0.0498	0.0189
60-69	0.5956	0.1524	0.2153	0.0289	0.0079
70-79	0.6808	0.1542	0.1327	0.0161	0.0161
Overdiagnosis rate 25%					
40-49	0.4452	0.1623	0.3147	0.0584	0.0195
50-59	0.4947	0.1963	0.2312	0.0564	0.0215
60-69	0.5417	0.1727	0.2440	0.0328	0.0089
70-79	0.6383	0.1748	0.1504	0.0183	0.0183
Triennial or quinquennial screening					
Overdiagnosis rate 5%					
40-49	0.4642	0.1659	0.2742	0.0646	0.0311
50-59	0.4838	0.1797	0.2308	0.0655	0.0402
60-69	0.5141	0.1689	0.2236	0.0515	0.0419
70-79	0.5695	0.1858	0.1573	0.0441	0.0434
Overdiagnosis rate 15%					
40-49	0.4012	0.1854	0.3065	0.0722	0.0347
50-59	0.4231	0.2008	0.2580	0.0732	0.0449
60-69	0.4569	0.1888	0.2499	0.0575	0.0468
70-79	0.5188	0.2076	0.1758	0.0493	0.0485
Overdiagnosis rate 25%					
40-49	0.3213	0.2101	0.3473	0.0819	0.0393
50-59	0.3461	0.2276	0.2924	0.0829	0.0509
60-69	0.3845	0.2140	0.2832	0.0652	0.0531
70-79	0.4547	0.2353	0.1992	0.0559	0.0549

¹ Distribution of stages at diagnosis of breast cancer for screen-detected cases with an overdiagnosis rate of 0%, same as Table S1.

Table S9: Linear regression model with dependent variable being the DCIS rate per 10⁵ women-year

Variables	Coefficient	SE	p-value
Intercept	0.4027	1.9045	0.838
Age group 50-69 years ¹	2.6302	1.9898	0.223
Age group ≥70 years ¹	2.8019	2.1886	0.236
Index of mammography usage (IMU) ²	31.1336	4.5488	<0.001

$R^2=0.87$

¹ The reference group for the variable age group is 40-49 years.

² Mammogram usage is an index that measures the intensity of mammography use per woman-year. It takes values between 0 and 1. If nobody is using mammography for early detection IMU=0. If every woman in the population receives an annual mammography, IMU=1. If every woman receives a biennial mammography, IMU=0.5.

Table S10: Cost-effectiveness and harm-benefit analysis. Lives extended (LE)¹

Cost-effectiveness analysis. Incremental cost per incremental LE, per 100,000 women at birth					
Strategy	LE	Cost (×10 ⁶ euros)	ΔLE	ΔCost (×10 ⁶ euros)	ICER euros/LE
Q5069-Q5069-Q5069-Q5074	146.6	123.8			
Q5069-Q5069-Q5074-Q5074	157.8	123.9	11.3	0.1	0.01
Q5069-Q5074-Q5074-Q5074	177.5	124.3	19.7	0.4	0.02
Q5074-Q5074-Q5074-Q5074	189.1	124.8	11.6	0.5	0.04
Q5074-Q5074-Q5074-Q4574	190.0	124.8	0.9	0.0	0.06
Q5074-Q5074-Q5074-T5074	191.5	125.0	1.6	0.1	0.09
Q5074-Q5074-Q5074-A5074	197.7	125.6	6.1	0.6	0.10
Q5074-Q5074-T5074-A5074	209.5	127.1	11.8	1.5	0.13
Q5074-Q5074-A5074-A5074	238.8	132.5	29.3	5.4	0.18
Q5074-T5074-A5074-A5074	258.7	136.7	19.9	4.2	0.21
Q5074-T5074-A5074-A4574	260.2	137.0	1.4	0.4	0.25
Q5074-A5074-A5074-A4574	309.5	152.8	49.3	15.7	0.32
T5074-A5074-A5074-A4574	320.4	156.6	10.9	3.9	0.35
T5074-A5074-A4574-A4574	327.2	159.7	6.8	3.1	0.45
T5074-A5074-A4574-A4074	328.2	160.2	1.0	0.5	0.46
A5074-A5074-A4574-A4074	354.9	175.4	26.7	15.2	0.57
A5074-A5074-A4074-A4074	359.7	179.3	4.8	3.9	0.80
A5074-A4574-A4074-A4074	370.7	188.1	11.0	8.8	0.80
A5074-A4074-A4074-A4074	378.6	198.9	7.8	10.8	1.38
A4574-A4074-A4074-A4074	384.3	207.3	5.7	8.4	1.46
A4074-A4074-A4074-A4074	388.4	217.4	4.1	10.2	2.47
Incremental false-positives (FP) per incremental LE, per 100,000 women at birth					
Strategy	LE	FP	ΔLE	ΔFP	ΔFP/ΔLE
Q5069-Q5069-Q5069-Q5069	144.2	12,995.4			
Q5069-Q5069-Q5069-Q5074	146.6	13,006.2	2.3	10.8	4.7
Q5069-Q5069-Q5074-Q5074	157.8	13,090.8	11.3	84.5	7.5
Q5069-Q5074-Q5074-Q5074	177.5	13,321.9	19.7	231.2	11.7
Q5074-Q5074-Q5074-Q5074	189.1	13,536.1	11.6	214.2	18.5
Q5074-Q5074-Q5074-T5074	191.5	13,607.4	2.4	71.3	29.1
Q5074-Q5074-Q5074-A5074	197.7	13,856.0	6.1	248.5	40.5
Q5074-Q5074-T5074-A5074	209.5	14,415.2	11.8	559.3	47.5
Q5074-Q5074-A5074-A5074	238.8	16,363.8	29.3	1,948.6	66.4
Q5074-T5074-A5074-A5074	258.7	17,893.3	19.9	1,529.5	76.7
Q5074-T5074-A5074-A4574	260.2	18,034.1	1.4	140.8	97.2
Q5074-A5074-A5074-A4574	309.5	23,362.9	49.3	5,328.8	108.1
T5074-A5074-A5074-A4574	320.4	24,780.0	10.9	1,417.1	129.5
A5074-A5074-A5074-A4574	347.1	29,717.0	26.7	4,937.1	184.8
A5074-A5074-A4574-A4574	353.9	31,031.0	6.8	1,313.9	194.4
A5074-A5074-A4574-A4074	354.9	31,240.1	1.0	209.2	201.0
A5074-A4574-A4574-A4074	365.9	35,176.2	11.0	3,936.1	358.2
A5074-A4574-A4074-A4074	370.7	36,994.8	4.8	1,818.6	376.2
A4574-A4574-A4074-A4074	376.5	40,844.6	5.7	3,849.9	670.7
A4574-A4074-A4074-A4074	384.3	46,107.9	7.8	5,263.3	671.2
A4074-A4074-A4074-A4074	388.4	51,155.8	4.1	5,047.9	1,224.1

Continued on next page...

...Continued from previous page.

Incremental overdiagnosis (ODX) per incremental LE, per 100,000 women at birth					
Strategy	LE	ODX	Δ LE	Δ ODX	Δ ODX/ Δ LE
Q5069-Q5069-Q5069-Q5069	144.2	187.2			
Q5069-Q5069-Q5069-Q5074	146.6	189.3	2.3	2.1	0.9
Q5069-Q5069-Q5074-Q5074	157.8	200.1	11.3	10.8	1.0
Q5069-Q5074-Q5074-Q5074	177.5	220.2	19.7	20.2	1.0
Q5074-Q5074-Q5074-Q5074	189.1	233.4	11.6	13.2	1.1
Q5074-Q5074-Q5074-Q4574	190.0	234.6	0.9	1.1	1.3
Q5074-Q5074-Q5074-Q4074	190.6	235.3	0.6	0.8	1.3
Q5074-Q5074-Q4574-Q4074	194.6	241.6	4.0	6.3	1.6
Q5074-Q5074-Q4074-Q4074	197.3	246.5	2.8	4.9	1.8
Q5074-Q5074-Q4574-A4574	203.1	257.6	5.8	11.0	1.9
Q5074-Q5074-A5074-A5074	238.8	326.9	35.7	69.3	1.9
Q5074-A5074-A5074-A5074	308.0	493.4	69.2	166.5	2.4
Q5074-A5074-A5074-A4574	309.5	497.2	1.4	3.8	2.6
Q5074-A5074-A5074-A4074	310.5	500.5	1.0	3.2	3.1
T5074-A5074-A5074-A4074	321.4	536.2	10.9	35.7	3.3
A5074-A5074-A5074-A4074	348.1	623.7	26.7	87.5	3.3
A5074-A5074-A4574-A4074	354.9	646.5	6.8	22.9	3.4
A5074-A5074-A4074-A4074	359.7	667.5	4.8	21.0	4.3
A5074-A4574-A4074-A4074	370.7	718.6	11.0	51.1	4.7
A5074-A4074-A4074-A4074	378.6	769.0	7.8	50.4	6.4
A4574-A4074-A4074-A4074	384.3	809.6	5.7	40.6	7.1
A4074-A4074-A4074-A4074	388.4	852.2	4.1	42.6	10.3

Incremental false-negatives (FN) per incremental LE, per 100,000 women at birth					
Strategy	LE	FN	Δ LE	Δ FN	Δ FN/ Δ LE
A5069-A5069-A5069-A5069	297.1	158.9			
A5074-A5074-A5074-A5074	345.7	180.9	48.6	22.0	0.5
A5074-A5074-A5074-A4574	347.1	182.8	1.4	1.8	1.3
A5074-A5074-A4574-A4574	353.9	191.3	6.8	8.6	1.3
A5074-A4574-A4574-A4574	364.9	205.2	11.0	13.9	1.3
A4574-A4574-A4574-A4574	370.6	212.5	5.7	7.3	1.3
A4574-A4574-A4574-A4074	371.6	215.4	1.0	2.9	2.7
A4074-A4074-A4074-A4074	388.4	261.6	16.8	46.2	2.8

¹ Data corresponding to a cohort of 100,000 women at birth assessed in the age-interval 40-79 years. All the absolute values have been discounted at an annual rate of 3%.

Table S11: Cost-effectiveness and harm-benefit analysis. Quality-adjusted life years (QALY)¹

Cost-effectiveness analysis. Incremental cost per incremental QALY, per 100,000 women at birth					
Schedule	QALY	Cost ($\times 10^6$ euros)	Δ QALY	Δ Cost ($\times 10^6$ euros)	Δ Cost/ Δ QALY euros/QALY
Q5069-Q5069-Q5069-Q5074	1,816.3	123.8			
Q5069-Q5069-Q5074-Q5074	1,869.5	123.9	53.2	0.1	0.001
Q5069-Q5069-Q5074-Q4574	1,892.0	123.9	22.5	0.0	0.002
Q5069-Q5074-Q5074-Q4574	1,983.9	124.3	91.9	0.4	0.004
Q5069-Q5074-Q4574-Q4574	2,084.5	124.9	100.6	0.6	0.006
Q5069-Q5074-Q4574-Q4074	2,101.8	125.0	17.3	0.1	0.006
Q5074-Q5074-Q4574-Q4074	2,154.7	125.5	52.9	0.5	0.010
Q5074-Q5074-Q4574-A4574	2,241.0	126.5	86.3	1.0	0.012
Q5074-Q4574-Q4574-A4574	2,393.3	128.5	152.3	1.9	0.013
Q5074-Q4574-Q4074-A4074	2,496.3	129.9	103.0	1.4	0.014
Q5074-Q4574-A4574-A4074	2,893.2	138.3	396.8	8.4	0.021
Q4574-Q4574-A4574-A4074	2,961.9	140.3	68.8	2.0	0.029
Q4574-Q4074-A4074-A4074	3,187.2	146.9	225.2	6.6	0.029
Q4569-A4569-A4069-A4074	3,682.2	165.6	495.1	18.7	0.038
Q4574-A4574-A4074-A4074	3,810.9	171.0	128.7	5.3	0.042
A5069-A4569-A4074-A4074	3,987.3	181.2	176.4	10.2	0.058
A5069-A4069-A4074-A4074	4,148.0	192.0	160.7	10.8	0.067
A5069-A4074-A4074-A4074	4,193.8	195.6	45.8	3.6	0.078
A4569-A4074-A4074-A4074	4,286.7	203.9	92.9	8.4	0.090
A4574-A4074-A4074-A4074	4,307.6	207.3	20.9	3.3	0.159
A4074-A4074-A4074-A4074	4,360.5	217.4	52.9	10.2	0.192

Continued on next page...

...Continued from previous page.

Incremental false-positive (FP) per incremental QALY, per 100,000 women at birth					
Schedule	QALY	FP	Δ QALY	Δ FP	Δ FP/ Δ QALY
Q5069-Q5069-Q5069-Q5069	1,805.3	12,995.4			
Q5069-Q5069-Q5069-Q5074	1,816.3	13,006.2	11.0	10.8	1.0
Q5069-Q5069-Q5069-Q4574	1,838.8	13,034.5	22.5	28.3	1.3
Q5069-Q5069-Q5074-Q4574	1,892.0	13,119.0	53.2	84.5	1.6
Q5069-Q5074-Q5074-Q4574	1,983.9	13,350.2	91.9	231.2	2.5
Q5074-Q5074-Q5074-Q4574	2,036.8	13,564.4	52.9	214.2	4.0
Q5074-Q5074-Q5074-A4574	2,140.4	13,996.8	103.6	432.4	4.2
Q5074-Q5074-Q4574-A4574	2,241.0	14,428.4	100.6	431.7	4.3
Q5074-Q5074-A5074-A4574	2,559.7	16,504.6	318.7	2,076.2	6.5
Q5074-Q5074-A5074-A4074	2,587.6	16,713.8	27.9	209.2	7.5
Q5074-Q5074-A4574-A4074	2,740.9	18,027.7	153.3	1,313.9	8.6
Q5074-Q4574-A4574-A4074	2,893.2	19,551.1	152.3	1,523.3	10.0
Q5069-A5069-A4569-A4074	3,271.0	23,362.4	377.8	3,811.4	10.1
Q5069-A5069-A4574-A4074	3,300.6	23,713.0	29.6	350.6	11.9
Q5074-A5074-A4574-A4074	3,399.3	24,886.0	98.8	1,173.0	11.9
Q5074-A5074-A4074-A4074	3,517.7	26,704.6	118.4	1,818.6	15.4
Q5074-A4574-A4074-A4074	3,742.2	30,640.6	224.5	3,936.1	17.5
A5069-A4569-A4074-A4074	3,987.3	35,147.7	245.1	4,507.0	18.4
A5069-A4574-A4074-A4074	4,033.1	36,106.5	45.8	958.8	20.9
A5069-A4074-A4074-A4074	4,193.8	41,369.7	160.6	5,263.3	32.8
A4569-A4074-A4074-A4074	4,286.7	45,219.6	92.9	3,849.9	41.5
A4574-A4074-A4074-A4074	4,307.6	46,107.9	20.9	888.3	42.5
A4074-A4074-A4074-A4074	4,360.5	51,155.8	52.9	5,047.9	95.3
Incremental overdiagnosis (ODX) per incremental QALY, per 100,000 women at birth					
Schedule	QALY	ODX	Δ QALY	Δ ODX	Δ ODX/ Δ QALY
Q5069-Q5069-Q5069-Q5069	1,805.3	187.2			
Q5069-Q5069-Q5069-Q4069	1,845.1	189.1	39.8	1.9	0.05
Q5069-Q5069-Q4569-Q4069	1,945.6	195.3	100.4	6.3	0.06
Q5069-Q5069-Q4069-Q4069	2,020.7	200.3	75.1	4.9	0.07
Q5069-Q4569-Q4069-Q4069	2,172.6	213.4	151.9	13.2	0.09
Q5069-Q4069-Q4069-Q4069	2,279.6	224.9	106.9	11.5	0.11
Q5069-Q4069-Q4069-A4069	2,398.3	241.8	118.7	16.9	0.14
Q4569-Q4069-Q4069-A4069	2,466.7	251.6	68.4	9.8	0.14
Q4569-Q4069-A4069-A4069	3,005.6	351.1	539.0	99.5	0.18
Q4069-Q4069-A4069-A4069	3,046.9	360.5	41.3	9.4	0.23
Q4569-A4569-A4069-A4069	3,675.8	518.5	628.8	158.0	0.25
Q4069-A4069-A4069-A4069	3,877.7	578.4	201.9	59.9	0.30
A5069-A4069-A4069-A4069	4,111.9	664.4	234.3	86.1	0.37
A4569-A4069-A4069-A4069	4,204.8	705.0	92.9	40.6	0.44
A4569-A4069-A4069-A4074	4,211.3	709.4	6.5	4.4	0.68
A4569-A4069-A4074-A4074	4,240.9	732.8	29.6	23.4	0.79
A4069-A4069-A4074-A4074	4,293.8	775.4	52.9	42.6	0.80
A4069-A4074-A4074-A4074	4,339.6	821.1	45.8	45.7	1.00
A4074-A4074-A4074-A4074	4,360.5	852.2	20.9	31.1	1.49
Incremental false-negatives (FN) per incremental QALY, per 100,000 women at birth					
Schedule	QALY	FN	Δ QALY	Δ FN	Δ FN/ Δ QALY
A5069-A5069-A5069-A5069	3,392.5	158.9			
A5069-A5069-A5069-A4569	3,427.3	160.7	34.8	1.8	0.05
A5069-A5069-A4569-A4569	3,580.5	169.3	153.3	8.6	0.06
A5069-A4569-A4569-A4569	3,805.0	183.2	224.5	13.9	0.06
A4569-A4569-A4569-A4569	3,897.9	190.5	92.9	7.3	0.08
A4569-A4569-A4569-A4069	3,925.8	193.3	27.9	2.9	0.10
A4569-A4569-A4069-A4069	4,044.2	206.7	118.4	13.3	0.11
A4569-A4069-A4069-A4069	4,204.8	228.3	160.7	21.6	0.13
A4569-A4069-A4069-A4074	4,211.3	229.5	6.5	1.2	0.19
A4569-A4069-A4074-A4074	4,240.9	235.4	29.6	5.9	0.20
A4069-A4069-A4074-A4074	4,293.8	246.7	52.9	11.3	0.21
A4069-A4074-A4074-A4074	4,339.6	256.5	45.8	9.8	0.21
A4074-A4074-A4074-A4074	4,360.5	261.6	20.9	5.1	0.24

¹ Data corresponding to a cohort of 100,000 women at birth assessed in the age-interval 40-79 years. All the absolute values have been discounted at an annual rate of 3%.

Table S12: Number of mammograms and detection rates for screen-detected and interval cases and program sensitivity by age groups. Invasive cancer (DCIS not included)

	Number of mammograms	Screen-detected		Interval cancer		Program sensitivity
	N	N	%	N	%	%
The INCA study						
Total	1,508,584	3,096	2.05	1,448	0.96	68.13
44-49	65,177	168	2.58	74	1.14	69.42
50-54	480,094	808	1.68	512	1.07	61.21
55-59	441,633	808	1.83	400	0.91	66.89
60-64	383,641	860	2.24	325	0.85	72.57
65-70	138,039	452	3.27	137	1.00	76.74
Mathematical model ¹ , biennial screening						
Total	1,186,681	3,021	2.55	1,399	1.18	68.35
45-49	235,524	297	1.26	192	0.82	60.74
50-54	233,102	460	1.97	251	1.08	64.70
55-59	229,681	555	2.42	257	1.12	68.35
60-64	225,542	669	2.97	352	1.56	65.52
65-70	262,832	1,040	3.96	347	1.32	74.98

¹ The results correspond to the fixed B4569 strategy.

Table S13: Distribution of stages at diagnosis of breast cancer

Age (years)	Stages (%)			
	I	II	III	IV
Screen-detected cases. INCA study				
44-49	64.7	23.5	9.8	2.0
50-59	59.0	31.3	8.9	0.8
60-69	64.9	26.3	8.0	0.8
Interval cases. INCA study				
44-49	30.6	45.2	21.0	3.2
50-59	27.5	43.8	24.8	4.0
60-69	28.5	46.1	20.9	4.5
Screen-detected cases. Model inputs ¹				
40-49	58.4	35.8	4.4	1.5
50-59	62.1	32.1	4.2	1.6
60-69	65.6	31.3	2.5	0.7
Interval cases. Model inputs ¹				
40-49	36.7	53.5	8.2	1.6
50-59	29.5	52.6	11.7	6.3
60-69	40.8	49.0	7.4	2.8

¹ The results correspond to the fixed B4569 strategy.

Table S14: Sensitivity analysis. Changes in lives extended^{1,2}

Overdiagnosis rate invasive tumors	Overdiagnosis rate (per 1,000) of DCIS	Changes in costs	LE N (%)	Cost N (%)	Overdiagnosis N (%)
0%	-	-	228.5 (1.1)	126.1 (-8.8)	63.6 (-42.5)
5%	-	-	222.8 (1.8)	126.4 (-8.9)	134.4 (-29.1)
-	0.10	-	209.5 (3.8)	126.6 (-8.7)	242.6 (-16.2)
15%	0.21	C ³	209.5 (3.8)	127.1 (-8.9)	275.9 (-20.6)
-	-	2 times C	209.5 (3.8)	219.4 (-5.6)	275.9 (-20.6)
-	-	5 times C	209.5 (3.8)	496.2 (-2.9)	275.9 (-20.6)
-	0.26	-	209.5 (3.8)	127.4 (-9.0)	291.1 (-22.1)
25%	-	-	192.5 (6.7)	128.1 (-9.0)	417.5 (-17.4)

¹ Data corresponding to a cohort of 100,000 women at birth assessed in the age-interval 40-79 years. All the absolute values have been discounted at an annual rate of 3%.

² Results obtained for the risk-based strategy Q5074-Q5074-T5074-A5074, and the percentages of change compared to the B5069 fixed strategy.

³ C: costs in the main analysis.

Table S15: Sensitivity analysis. Changes in QALY^{1,2}

Overdiagnosis rate invasive tumors	Overdiagnosis rate (per 1,000) of DCIS	Changes in costs	Changes in disutility by FP	QALY N (%)	Cost N (%)	Overdiagnosis N (%)
0%	-	-	-	2616.3 (-1.0)	127.4 (-7.8)	68.3 (-38.2)
5%	-	-	-	2542.4 (-0.3)	127.7 (-7.9)	132.4 (-30.1)
-	0.10	-	-	2371.2 (1.4)	127.9 (-7.8)	224.9 (-22.3)
-	-	-	0 times D ⁴	2492.5 (-0.5)	128.4 (-8.0)	260.7 (-25.0)
15%	0.21	C ³	D	2368.3 (1.5)	128.4 (-8.0)	260.7 (-25.0)
-	-	-	2 times D	2244.1 (3.8)	128.4 (-8.0)	260.7 (-25.0)
-	-	2 times C	-	2368.3 (1.5)	220.7 (-5.0)	260.7 (-25.0)
-	-	5 times C	-	2368.3 (1.5)	497.4 (-2.6)	260.7 (-25.0)
-	0.26	-	-	2367.0 (1.5)	128.7 (-8.1)	277.0 (-25.9)
25%	-	-	-	2147.4 (4.3)	129.3 (-8.1)	389.0 (-23.1)

¹ Data corresponding to a cohort of 100,000 women at birth assessed in the age-interval 40-79 years. All the absolute values have been discounted at an annual rate of 3%.

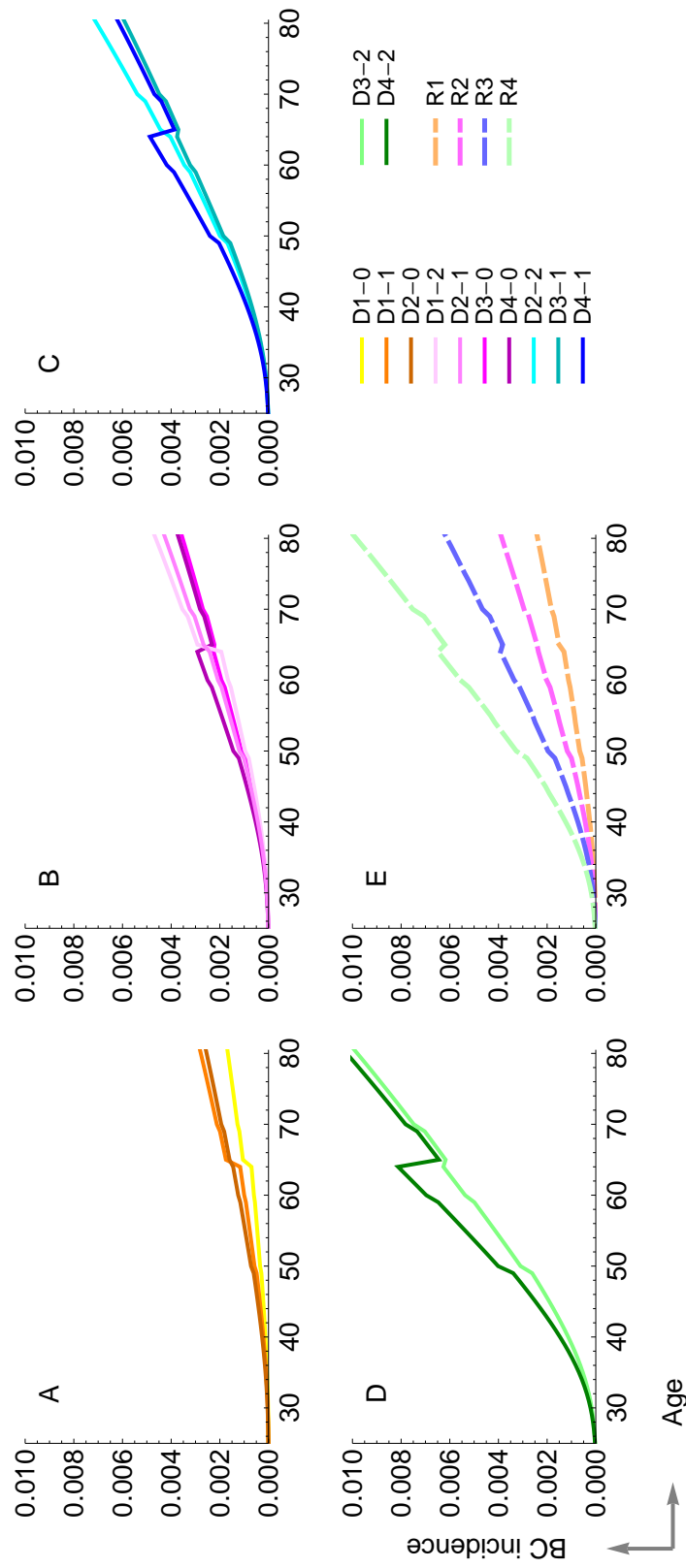
² Results obtained for the risk-based strategy Q5069-Q4574-A4074, and the percentages of change compared to the B5069 fixed strategy.

³ C: costs in the main analysis.

⁴ D: disutility in the main analysis.

Figures

Figure S1: Incidence curves for twelve risk profiles grouped by risk level: (A) Low Risk, (B) Medium-Low Risk, (C) Medium-High Risk, and (D) High Risk. Graphic (E) shows the smoothed incidence rates for each risk group



Based on Schousboe *et al.* [10], Tice *et al.* [9], and the Risk Estimation Dataset of the Breast Cancer Surveillance Consortium (BCSC) [11].

Figure S2: Observed and smoothed DCIS rates over time in Catalonia (1983-2008)

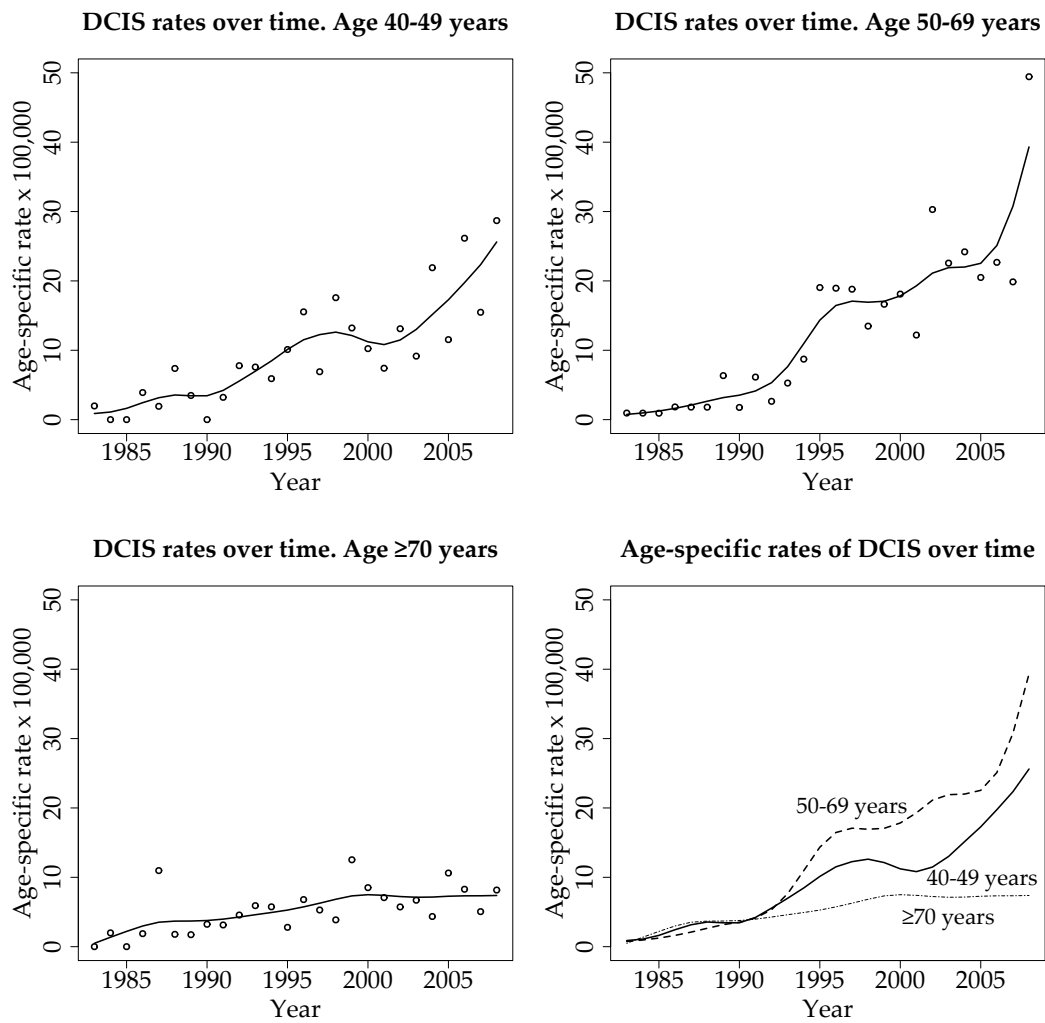


Figure S3: Index of mammography use (IMU) and smoothed DCIS rates over time in Catalonia (1983-2008)

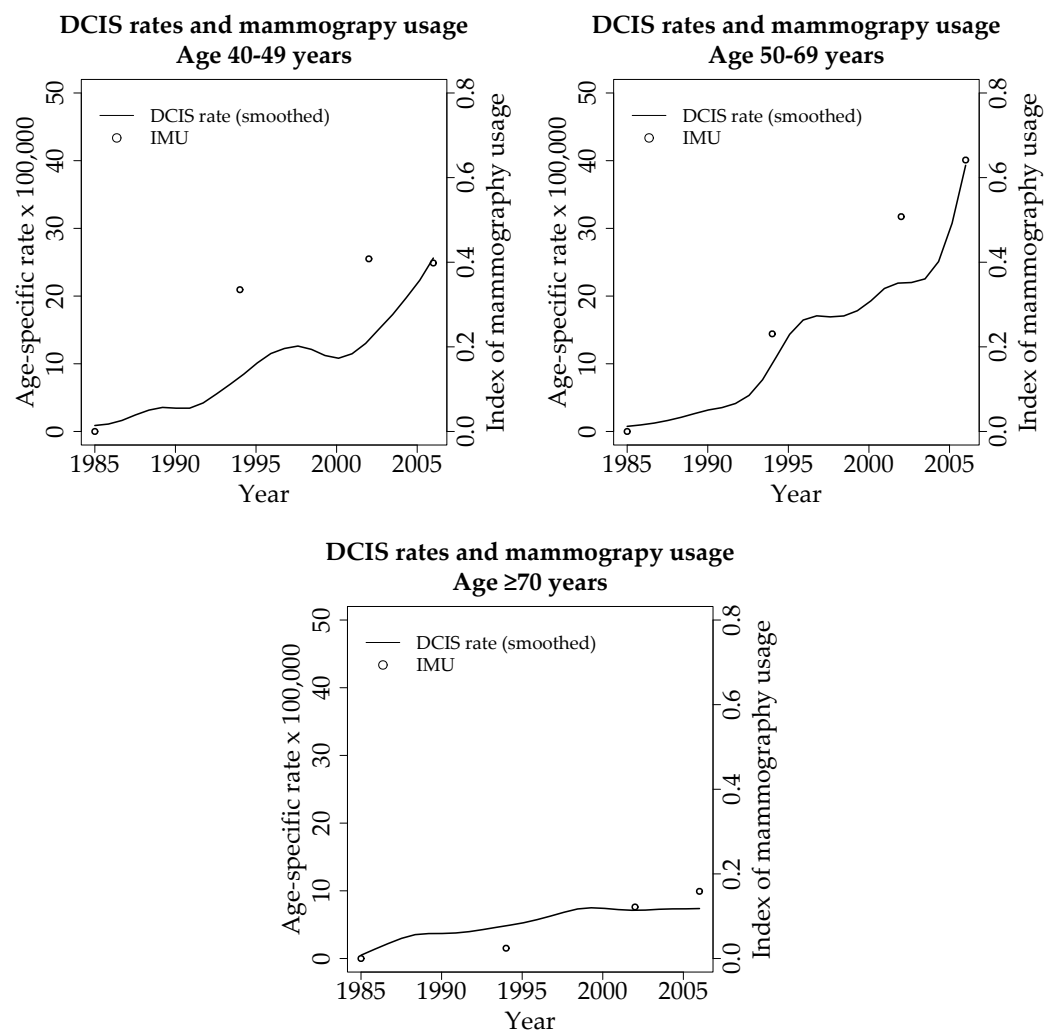
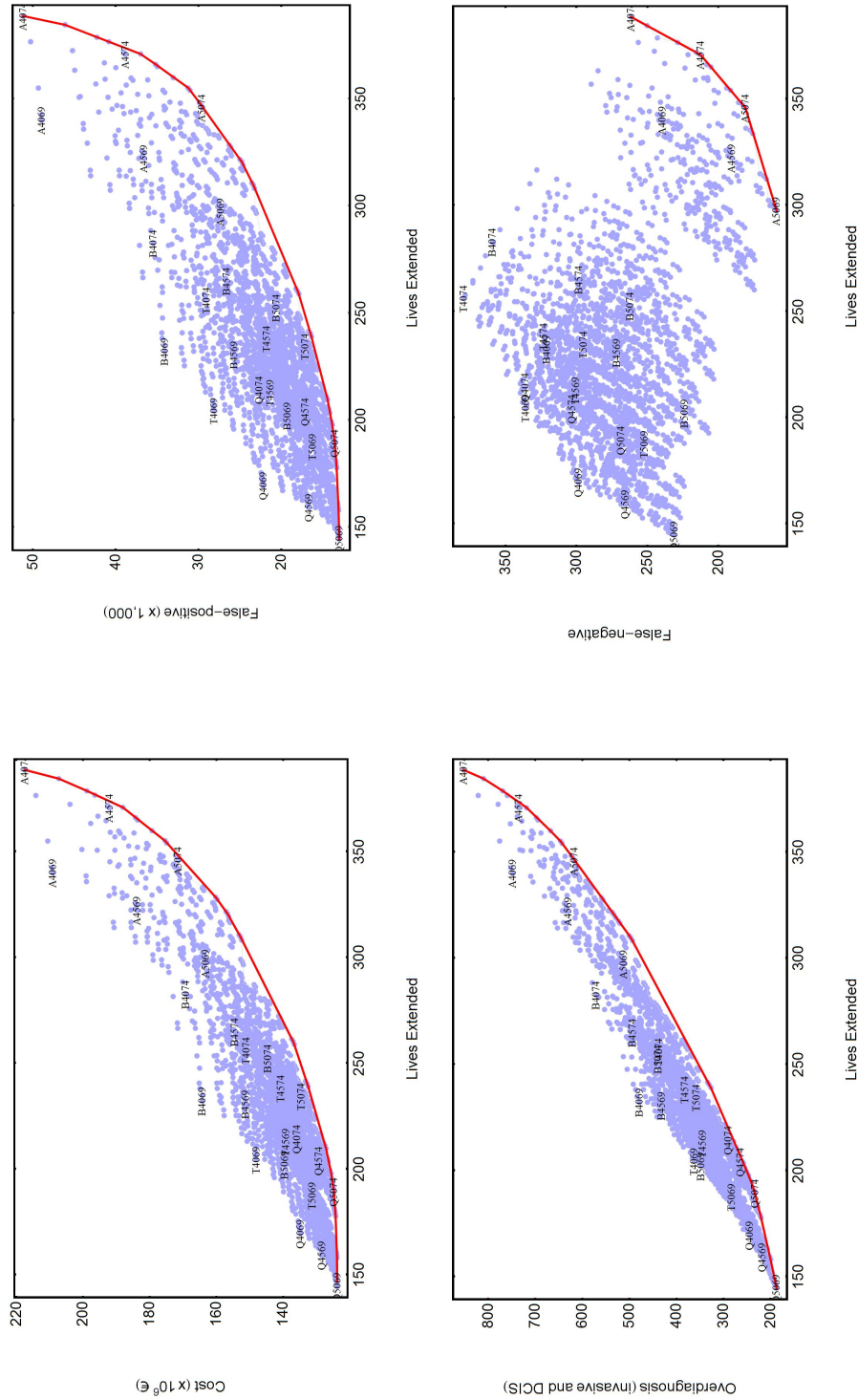
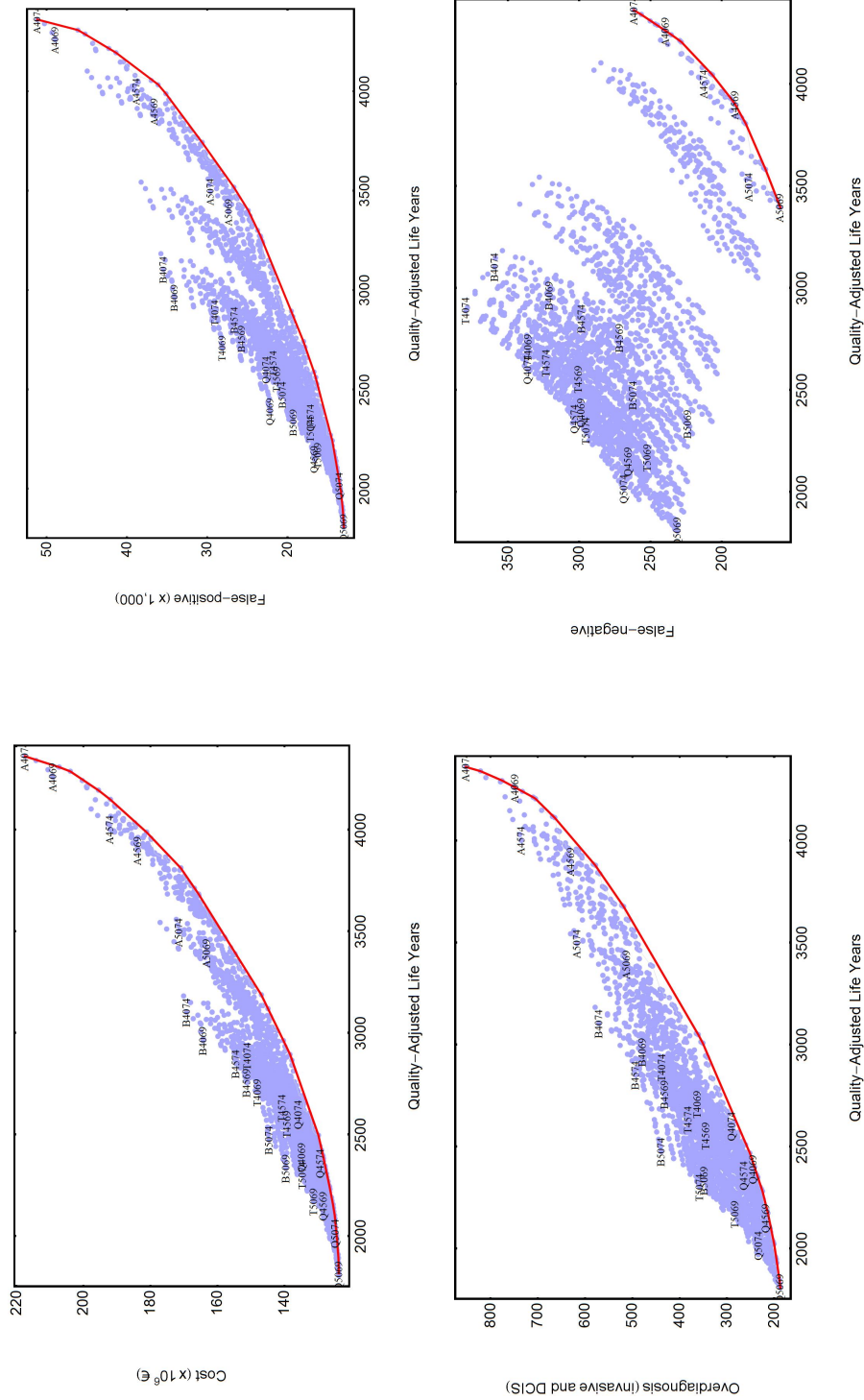


Figure S4: Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies, with uniform strategies marked. Effect measured in lives extended



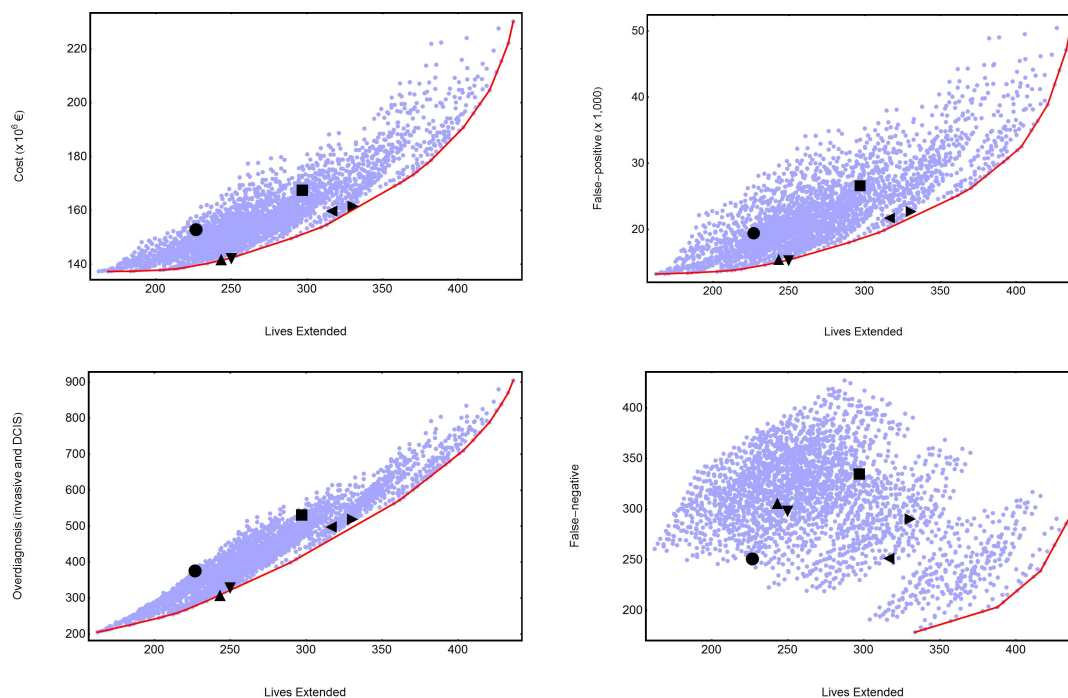
Results obtained with an annual discount of 3%. Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial. The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam.

Figure S5: Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies, with uniform strategies marked. Effect measured in quality-adjusted life years



Results obtained with an annual discount of 3%. Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial. The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam.

Figure S6: Sensitivity analysis of a change in the risk groups distribution. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in lives extended



●: uniform B5069; ■: uniform B4574

▲: risk-based Q5074-Q5074-Q4574-A4574; ▼: risk-based Q5074-Q5074-T5074-A5074

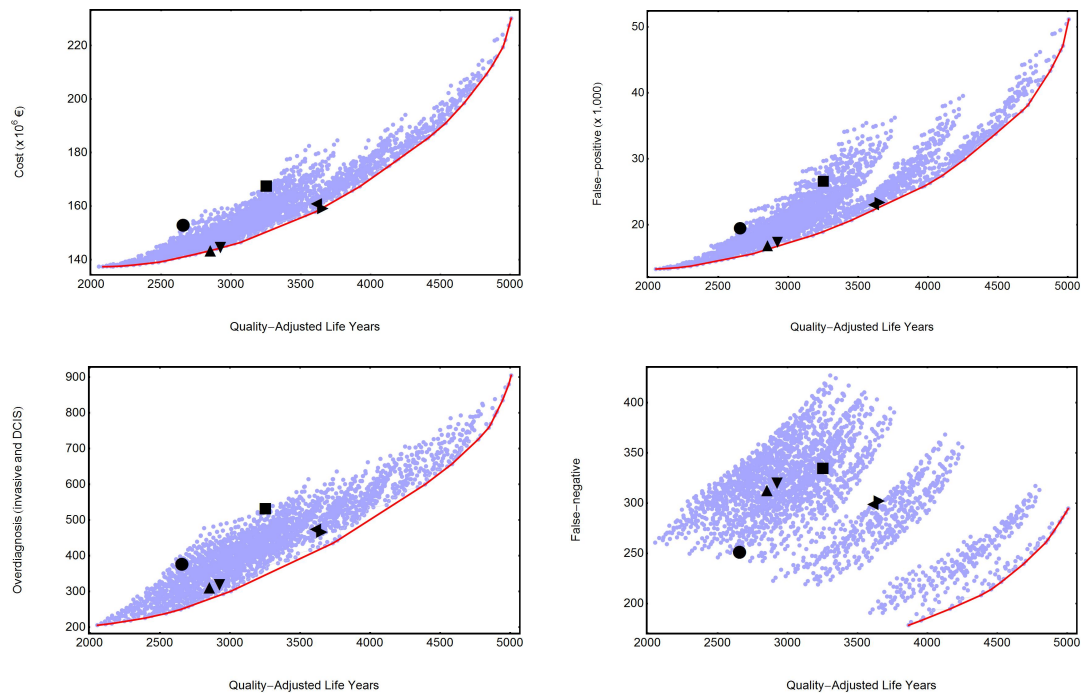
◄: risk-based T5069-B5074-A5074-A5074; ►: risk-based T5074-T5074-A4574-A4574

Results obtained with an annual discount of 3%.

Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial.

The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam. In the risk-based strategies, the four strings correspond to the Low, Medium-Low, Medium-High and High risk groups, respectively.

Figure S7: Sensitivity analysis of a change in the risk groups distribution. Cost-effectiveness and harm-benefit analyses for 2,625 early detection strategies. Effect measured in quality-adjusted life years



●: uniform B5069; ■: uniform B4574

▲: risk-based Q5069-Q4574-Q4574-A4574; ▼: risk-based Q5069-Q4574-Q4574-A4074

◄: risk-based Q5074-Q5074-A4074-A4074; ►: risk-based Q4574-Q4574-A4574-A4074

Results obtained with an annual discount of 3%.

Exams periodicities: A = annual, B = biennial, T = triennial, Q = quinquennial.

The first two numbers refer to the age at starting the exams and the last two numbers refer to the age at the last exam. In the risk-based strategies, the four strings correspond to the Low, Medium-Low, Medium-High and High risk groups, respectively.

References

1. Lee S, Zelen M (1998) Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *J Am Stat Assoc* (93):1271-1281.
2. Lee S, Zelen M (2006) A stochastic model for predicting the mortality of breast cancer. *J Natl Cancer Inst Monogr* (36):79-86.
3. Lee SJ, Zelen M (2008) Mortality modeling of early detection programs. *Biometrics* 64:386-395.
4. Vilapriño E, Rue M, Marcos-Gragera R, Martínez-Alonso M (2009) Estimation of age- and stage-specific Catalan breast cancer survival functions using US and Catalan survival data. *BMC Cancer* 9:98
5. Mariotto A, Feuer EJ, Harlan LC, Wun LM, Johnson KA, et al. (2002) Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975-1999. *J Natl Cancer Inst* 94:1626-1634.
6. Chia SK, Speers CH, D'yachkova Y, Kang A, Malfair-Taylor S, et al. (2007) The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. *Cancer* 110:973-979.
7. Martínez-Alonso M, Vilapriño E, Marcos-Gragera R, Rue M (2010) Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* 12:R58.
8. American College of Radiology (2003) The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). American College of Radiology, Reston (VA).
9. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, et al. (2008) Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 148:337-347.
10. Schousboe JT, Kerlikowske K, Loh A, Cummings SR (2011) Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med* 155:10-20.
11. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, et al. (2006) Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 98:1204-1214.
12. Lidgren M, Wilking N, Jonsson B, Rehnberg C (2007) Health related quality of life in different states of breast cancer. *Qual Life Res* 16:1073-1081.

13. Rue M, Carles M, Vilaprinyo E, Martinez-Alonso M, Espinas JA, et al. (2008) Dissemination of periodic mammography and patterns of use, by birth cohort, in Catalonia (Spain). *BMC Cancer* 8:336.
14. Perez MJ, Gregori A, Carles M, Gispert R, Martinez-Alonso M, et al. (2010) The evolution of breast cancer mortality and the dissemination of mammography in Catalonia: an analysis by health region. *Rev Esp Salud Publica* 84: 691-703.

3.2 Mètodes i resultats de l'objectiu 2

En aquesta secció es presenten els mètodes utilitzats i els resultats obtinguts per donar resposta al segon dels objectius platejats en aquesta tesi. Tota la metodologia i els resultats estan continguts en el manuscrit original Armero *et al.* (Statistics in Medicine 2016), la informació del qual es detalla a continuació, junt amb el resum traduït al català.

Títol: Bayesian joint ordinal and survival modeling for breast cancer risk assessment

Autors: Armero C, Forné C, Rué M, Forte A, Perpiñán H, Gómez G, Baré M

Revista: Stat Med. 2016 Dec 10;35(28):5267-5282

Factor d'impacte de l'any de la publicació: 1,861 (percentil 84,3 a *Statistics & Probability*)

DOI: 10.1002/sim.7065

Resum:

Proposem un model conjunt per analitzar l'estructura i la intensitat de l'associació entre les mesures longitudinals d'un marcador ordinal i el temps fins a un esdeveniment d'interès. El procés longitudinal es defineix en termes d'un model lògic acumulatiu d'odds proporcionals. El temps fins a l'esdeveniment es modelitza a través d'un model de riscos proporcionals truncat per l'esquerra, que incorpora la informació del marcador longitudinal i de les covariables basals. Els processos longitudinals i de supervivència estan connectats mitjançant un vector comú d'efectes aleatoris.

Les inferències es discuteixen sota l'enfocament bayesià i inclouen la distribució posterior de les probabilitats associades a cada categoria longitudinal i l'avaluació de l'impacte de les covariables basals i del marcador longitudinal a la funció de risc. La flexibilitat que proporciona el model conjunt permet estimar dinàmicament la probabilitat individual lliure de l'esdeveniment i predir els valors futurs del marcador longitudinal.

El model s'aplica a l'avaluació del risc de càncer de mama en dones que assisteixen a un programa de cribratge poblacional. El marcador ordinal longitudinal és la densitat mamogràfica mesurada amb l'escala BI-RADS en exàmens de cribratge biennal.

Bayesian joint ordinal and survival modeling for breast cancer risk assessment

C. Armero,^{a,*†} C. Forné,^{b,c} M. Rué,^{b,d} A. Forte,^a H. Perpiñán,^{a,e}
G. Gómez^f and M. Baré^g

We propose a joint model to analyze the structure and intensity of the association between longitudinal measurements of an ordinal marker and time to a relevant event. The longitudinal process is defined in terms of a proportional-odds cumulative logit model. Time-to-event is modeled through a left-truncated proportional-hazards model, which incorporates information of the longitudinal marker as well as baseline covariates. Both longitudinal and survival processes are connected by means of a common vector of random effects.

General inferences are discussed under the Bayesian approach and include the posterior distribution of the probabilities associated to each longitudinal category and the assessment of the impact of the baseline covariates and the longitudinal marker on the hazard function. The flexibility provided by the joint model makes possible to dynamically estimate individual event-free probabilities and predict future longitudinal marker values.

The model is applied to the assessment of breast cancer risk in women attending a population-based screening program. The longitudinal ordinal marker is mammographic breast density measured with the Breast Imaging Reporting and Data System (BI-RADS) scale in biennial screening exams. © 2016 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Keywords: BI-RADS scale; Latent process; Left-truncated proportional-hazards model; Proportional-odds cumulative logit model

1. Introduction

The current evidence on benefits and harms supports the personalization of screening as a crucial step to improve early detection of breast cancer [1, 2]. A number of risk models were designed to measure the individual probability of developing breast cancer [3–5]. In the context of individualized breast cancer screening, the utility of these risk models has been questioned because of their low discrimination power [6]. The inclusion of a baseline measure of breast density – a characteristic of the breast tissue – in the risk models improved the accuracy of the breast cancer risk estimate [7–10].

Several studies have shown that high breast density is associated with increased breast cancer risk [7, 11–14], with risk estimates in the range four-fold to six-fold for women with very high breast density compared with women with low breast density [11, 12]. Other studies have examined whether changes in breast density over time are associated to changes in breast cancer risk [15–21] and have suggested that

^aDepartment of Statistics and Operational Research, Universitat de València, Doctor Moliner, 50, 46100 Burjassot, Spain

^bDepartment of Basic Medical Sciences, Universitat de Lleida-IRBLleida, Avda. Rovira Roure, 80, 25198 Lleida, Spain

^cOblikue Consulting, Barcelona, Spain

^dHealth Services Research Network in Chronic Diseases (REDISSEC), Spain

^eFundación para el Fomento de la Investigación Sanitaria y Biomédica (FISABIO), Generalitat Valenciana, Spain

^fDepartment of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain

^gClinical Epidemiology and Cancer Screening, Corporació Sanitària Parc Taulí-UAB, Sabadell, Parc Taulí s/n, 08208 Sabadell, Spain

*Correspondence to: Carmen Armero, Department of Statistics and Operational Research, Universitat de València, Doctor Moliner, 50, 46100 - Burjassot, Spain.

†E-mail: Carmen.Armero@uv.es

The copyright line for this article was changed on 26 September 2016 after original online publication.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

monitoring changes in breast density could help to identify women at greater risk of disease. In most of the cases, the statistical methods used did not account for relevant characteristics of prospective studies like non-ignorable dropout mechanisms or internal time-dependent covariates [22].

Joint modeling of longitudinal and time-to-event data is an increasingly productive area of statistical research that assesses the association between longitudinal and survival processes. It enhances longitudinal modeling by allowing for the inclusion of non-ignorable dropout mechanisms and survival modeling by the inclusion of internal time-dependent covariates [22]. Joint models were introduced during the 90s [23–25] and since then, have been applied to a great variety of studies in epidemiological and biomedical areas. Shared-parameter models are a type of joint models where the longitudinal and time-to-event processes are connected by means of a common set of subject-specific random effects. These models make possible to quantify both the population and individual effects of the underlying longitudinal outcome on the risk of an event and obtain individualized time-dynamic predictions. Recently, Rizopoulos proposed an overview of the theory and applications of joint modeling [26] and developed the JM [27] and JMbayes [28] R packages for the frequentist and Bayesian shared-effects' approaches, respectively. Serrat *et al.* illustrate the application of both statistical approaches to joint modeling longitudinal measures of prostate specific antigen and prostate cancer detection in men participating in a screening trial [29].

When longitudinal outcomes are ordinal, joint models become more complex. Different approaches, that use constraints in the probabilities of the categorical outcomes or the discretization of a continuous latent variable, have been proposed [30–33]. The non-linear and longitudinal nature of the data produce a complex likelihood function, difficult to maximize under the frequentist paradigm. This could be a reason why the standard software for joint models does not include longitudinal ordinal variables yet. Some relevant works on the subject use the frequentist [34–36] and Bayesian [33, 37, 38] approaches, respectively.

The objective of this paper is to propose a Bayesian joint model for assessing the structure and intensity of the association between longitudinal measures of an ordinal marker and a time-to-event outcome. In particular, we use a proportional-odds cumulative logit model [30] for the ordinal measurements and a proportional hazard model with left-truncation for the time to an event of interest. We have applied the model to analyze the risk of breast cancer in women attending a population-based screening program with regard to repeated measurements of mammographic breast density.

Section 2 presents a description of the motivating dataset. Section 3 formulates the joint model and discusses general inferences for (1) dynamic probabilities associated to the different ordinal categories, (2) the impact of baseline covariates and the longitudinal marker on the hazard function, (3) dynamic estimation of survival probabilities, and (4) prediction of future longitudinal outcomes. Section 4 applies the model developed in Section 3 to study age at diagnosis of breast cancer in women who participate in a population-based screening program. Finally, Section 5 contains a discussion and some conclusions.

2. Motivating data

2.1. Study design and study population

This is an observational prospective study including 13,760 women that participated for the first time in the breast cancer early-detection program in the Vallès Occidental Est (BCEDP-VOE) area in Catalonia (Spain), between October 1995 and June 1998. The BCEDP-VOE invites women aged 50–69 years for biennial mammographic exams. At study entry, the participants were 50–70 years old and did not have a personal history of breast cancer. They were followed for vital status or possible diagnosis of breast cancer until December 2013 [39–41].

Of the initial 13,760 women, we excluded seven without follow-up data, as well as 38 women who were diagnosed with breast cancer and nine who died within six months of baseline. Twenty-one women were also excluded for not having breast density measurements within the 50–70 age interval. We analyzed invasive breast cancer and ductal carcinoma in situ diagnosed during follow-up. The final sample included 13,685 women, with 431 diagnosed with breast cancer.

2.2. Variables and data description

At the first mammographic exam, the study participants answered a questionnaire that included information on family history of breast cancer, prior breast procedures, age at menarche, age at first birth, and menopausal status. Family history refers to absence/presence of first-degree relatives with breast cancer. Prior breast procedures included breast biopsy, fine needle aspiration, cyst aspiration, breast reconstruction, lumpectomy, and surgical treatment.

Table I. Baseline risk factors according to breast cancer diagnosis status at the end of follow-up.

	No breast cancer <i>N</i> = 13,254 (%)		Breast cancer <i>N</i> = 431 (%)	
Family history of breast cancer				
No	12,539	(94.8)	388	(90.2)
Yes	686	(5.2)	42	(9.8)
Prior breast procedures				
No	12,318	(92.9)	374	(86.8)
Yes	936	(7.1)	57	(13.2)
Breast density at first examination (baseline breast density)				
a: Almost entirely fatty	2959	(23.4)	56	(13.9)
b: Scattered fibroglandular densities	5353	(42.3)	138	(34.2)
c: Heterogeneously dense	2301	(18.2)	103	(25.6)
d: Extremely dense	2037	(16.1)	106	(26.3)
Breast density at last examination ^a (women with at least two examinations)				
a: Almost entirely fatty	2284	(18.1)	35	(9.4)
b: Scattered fibroglandular densities	7957	(63.0)	201	(54.0)
c: Heterogeneously dense	1475	(11.7)	71	(19.1)
d: Extremely dense	919	(7.3)	65	(17.5)

^aIf breast cancer was diagnosed within 6 months following the last mammography, the last breast density considered was the previous one, whenever it was not coincident with the baseline measure.

Breast density is a characteristic of the breast tissue that is reflected in mammograms. Breasts are considered dense if the connective and epithelial tissues predominate over the fatty tissue. At all mammographic exams, breast density was rated and recorded according to the BI-RADS system [40, 42] that categorizes breast density in four groups: a, almost entirely fatty (low density); b, scattered fibroglandular densities (medium density); c, heterogeneously dense (high density); and d, extremely dense (very high density). This longitudinal breast density data is a remarkable and unique characteristic of the BCEDP-VOE among the breast cancer screening programs in Spain. Breast density measures within 6 months before breast cancer diagnosis could be affected by the presence of preclinical breast cancer; therefore, they were excluded. The mammographic exams performed before age 50 or after age 70 were excluded in order to avoid sample biases. A total of 81,621 measures of breast density were included in the longitudinal analysis, with median [range] 4 [1 to 9] and 6 [1 to 15] measures for women with or without breast cancer diagnosis, respectively.

We considered that the time origin for the event of interest (diagnosis of breast cancer) was age 50 years, the lower limit of the screening age interval. We defined the time-to-event as the time elapsed from the origin to diagnosis of breast cancer. For women without a breast cancer diagnosis at the study end, the censoring time was obtained as the minimum of time to death and time to the last screening exam plus 2.5 years that correspond to the active follow-up for cancer identification. It is important to remark that women who entered the program over 50 years had delayed entry times that may induce length biased sampling or left truncation.

Among 13,685 women aged 50 years and older, 431 developed breast cancer – 336 invasive cancers and 95 ductal carcinoma in situ –, and 513 died within 2.5 years of the last mammogram. Median follow-up was 12.7 years for women without breast cancer and 8.2 years for women with breast cancer. Table I shows the baseline characteristics of women and the breast density measurements at first and last examination according to their breast cancer diagnosis status at the end of follow-up. High breast density categories were more prevalent among women with breast cancer, in both the first and the last mammogram. Furthermore, between the first and last exams, the prevalence of low density categories increased, as described in the literature.

To illustrate the longitudinal breast density measurements, we randomly selected eight women without and eight women with cancer (Figure 1). A high variability of the breast density trajectories can be observed: some women experience an increase of breast density, while others remain stable, fluctuate, or experience a decrease. The plots show the biennial periodicity of the screening exams, as well as the unbalanced number of measures between women, because of different reasons. Not all women entered

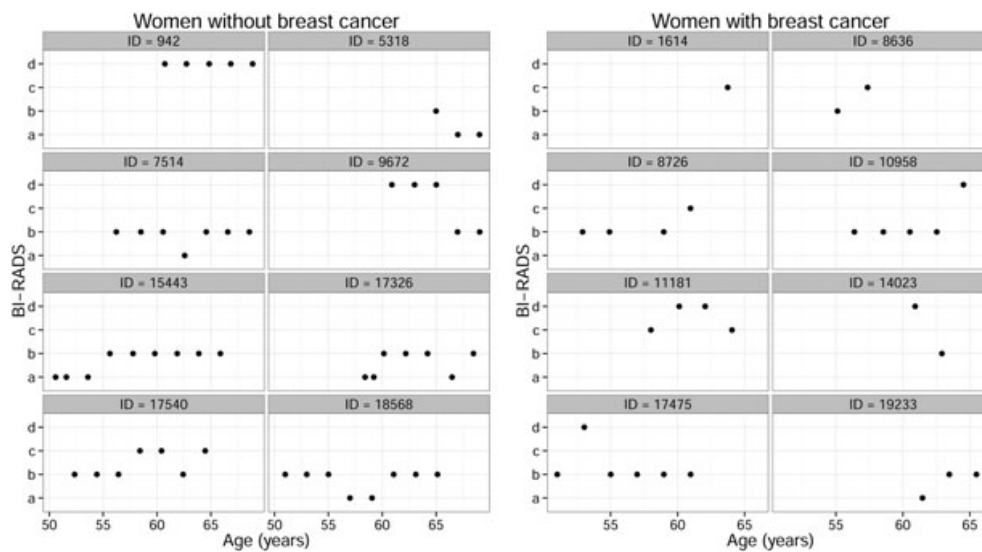


Figure 1. Subject-specific profiles of BI-RADS measures for 16 randomly selected women. The left panel corresponds to eight women without breast cancer, and the right panel corresponds to eight women with breast cancer.

the study at the same age, not all the scheduled screening exams were taken, or not always breast density was rated.

3. The joint model

We propose a model with two submodels: (1) a proportional-odds cumulative logit model for the longitudinal ordinal measurements based on the idea of a continuous latent variable [30, 33] and (2) a left-truncated proportional hazard model for the time-to-event, which incorporates information from the longitudinal process. Both processes are connected through a shared vector of random effects, which, in the presence of covariates and parameters, endows them with conditional independence [26].

Let $\{D_1, D_2, \dots, D_K\}$ denote the set of ordinal categories and y_{ij} the longitudinal category of individual i , $i = 1, \dots, n$, at time t_{ij} , $j = 1, \dots, n_i$. We assumed an underlying continuous latent variable y_{ij}^* that determines the ordinal category of individual i at time t_{ij} . This latent variable has no interest *per se*, but it is useful for motivating and interpreting the cumulative logit model. The relationship between y_{ij} and y_{ij}^* is stated as

$$y_{ij} = D_k \Leftrightarrow y_{ij}^* \in (\gamma_{k-1}, \gamma_k], \quad k = 1, \dots, K,$$

where $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{K-1} < \gamma_K = \infty$ are unknown cut-points. We assumed a logistic distribution for y_{ij}^* , $\text{Lo}(m_{ij}, s = 1)$, with location parameter m_{ij} (mean) and a common scale parameter $s = 1$ for achieving identifiability. The choice of that distribution implies a logit link for the cumulative probabilities

$$q_{ijk} = P(y_{ij} > D_k) = P(y_{ij}^* > \gamma_k) = \frac{1}{1 + \exp(\gamma_k - m_{ij})}, \quad (1)$$

and therefore,

$$\text{logit } q_{ijk} = \log \left(\frac{q_{ijk}}{1 - q_{ijk}} \right) = m_{ij} - \gamma_k.$$

Despite the fact that $s = 1$ in the logistic distribution of the latent variable, the model is overparameterized (any set of k probabilities can be obtained increasing the cut-points in the same quantity). To obtain an identifiable model, we arbitrarily introduced a reference point on the latent scale, in particular $\gamma_{K/2} = 0$ if K is even and $\gamma_{(K-1)/2} = 0$ or $\gamma_{(K+1)/2} = 0$ if K is odd.

We considered a mixed-effects model to describe the subject-specific time trajectories of the latent variable

$$y_{ij}^* = m_{ij} + \epsilon_{ij} = \mathbf{x}_{ij}^{(l)'} \boldsymbol{\beta} + \mathbf{z}_{ij}' \mathbf{b}_i + \epsilon_{ij}, \quad (2)$$

where $\mathbf{x}_{ij}^{(l)}$ is a P dimensional vector of covariates relevant to the longitudinal process, as indicated by superscript (l) , for individual i at time t_{ij} with regression coefficients (populational) β ; \mathbf{z}_{ij} the vector of explanatory variables attached to the vector of random effects \mathbf{b}_i for the i th individual at time t_{ij} ; and ϵ_{ij} an error term for the i th individual at time t_{ij} , modeled in terms of a logistic distribution, $\text{Lo}(0, 1)$. The random effects $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)^T$ are conditionally i.i.d. (given the hyperparameter vector ϕ) with $(\mathbf{b}_i | \phi) \sim f(\mathbf{b}_i | \phi)$.

Let T_i , $i = 1, \dots, n$, be the observed event time for the i th subject, obtained as the minimum between the true failure time, T_i^* , and the right-censoring time, C_i , $T_i = \min(T_i^*, C_i)$. The event indicator $\delta_i = I(T_i^* \leq C_i)$ takes the value 1 if the observed time corresponds to a true event time, and 0 otherwise. In addition, event times corresponding to individuals who enter the study at delayed entry times introduce left-truncation, thus defining the subsequent hazard function as zero in the period before the entrance of the individual to the system [43]. In particular, we consider the hazard function of T_i^* in terms of a left-truncated relative risk regression model

$$h_i(t) = h_0(t) \exp \left\{ \mathbf{x}_i^{(s)'} \boldsymbol{\eta} + \alpha m_{it} \right\}, \quad t > a_i, \quad (3)$$

and zero otherwise, where $h_0(t)$ is the baseline risk function; $\mathbf{x}_i^{(s)}$ represents the vector of baseline covariates relevant to the survival process, as indicated by superscript (s) , with associated coefficients $\boldsymbol{\eta}$; α assesses the effect of the longitudinal marker of subject i on the event of interest in terms of the latent variable mean; a_i is the delayed entry time of individual i . It is important to comment that left-truncated data will add computational complexity to the modeling because the likelihood function corresponding to this type of data will incorporate conditional probabilities that contain the information that the individual is alive in the period between their theoretical entrance at time zero and their real entrance to the system.

To complete the Bayesian model, it is necessary to elicit a prior distribution, $\pi(\boldsymbol{\theta})$ for all the unknown parameters and hyperparameters of the joint model $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\eta}, \alpha, \boldsymbol{\gamma}, \boldsymbol{\phi})^T$. Our joint model contains parameters and hyperparameters, $\boldsymbol{\theta}$, and random effects \mathbf{b} . From a Bayesian perspective, $\pi(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D})$, where \mathcal{D} represents all the data collected from the longitudinal and the survival processes, is the joint posterior distribution of the parameters, hyperparameters, and random effects, which can be obtained by the hierarchical modeling

$$\pi(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D}) = L(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D}) f(\mathbf{b} | \boldsymbol{\phi}) \pi(\boldsymbol{\theta}). \quad (4)$$

$L(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D})$ is the likelihood function of $\boldsymbol{\theta}$ and \mathbf{b} for data \mathcal{D} , $f(\mathbf{b} | \boldsymbol{\phi})$ the distribution of the random effects \mathbf{b} given $\boldsymbol{\phi}$ introduced before, and $\pi(\boldsymbol{\theta})$ the prior distribution for $\boldsymbol{\theta}$. Markov Chain Monte Carlo (MCMC) simulation methods allow to obtain an approximated random sample from the posterior $\pi(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D})$, which is the key element and the starting point of all relevant inferences.

Finally, it is worth noting that when inference is carried out under the Bayesian formulation, the shared joint model will induce conditional independence between the longitudinal and the survival processes given not only the random effects and covariates but also given all the parameters and hyperparameters in the model, as a result of its stochastic role in Bayesian inference.

3.1. Dynamic probabilities associated to ordinal categories

From expression (1), the probability distribution of the ordinal marker y_{it} for individual i at time t can be computed in terms of the logistic distribution of their latent variable y_{it}^* as

$$P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \boldsymbol{\theta}, \mathbf{b}_i) = P(y_{it}^* \in (\gamma_{k-1}, \gamma_k) | \mathbf{x}_{it}^{(l)}, \boldsymbol{\theta}, \mathbf{b}_i), \quad k = 1, 2, \dots, K. \quad (5)$$

These probabilities depend on $\boldsymbol{\theta}$, \mathbf{b}_i , and the relevant covariates associated to that individual. Consequently, we could use the posterior marginal distribution $\pi(\boldsymbol{\theta}, \mathbf{b}_i | \mathcal{D})$ for computing the posterior distribution, $\pi(P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \boldsymbol{\theta}, \mathbf{b}_i) | \mathcal{D})$, of all the relevant dynamic probabilities for each individual in the study.

A complementary and overall perspective of the temporal evolution of the different categories of the ordinal marker is based on the marginal distribution

$$P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \boldsymbol{\theta}) = \int P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{b}_i | \boldsymbol{\phi}) d\mathbf{b}_i, \quad (6)$$

which is computed by integrating out the random effects of the conditional distribution (5). This distribution only depends on θ . It can be interpreted as the time-specific population distribution of the longitudinal marker for a generic individual of the population with covariate values $\mathbf{x}_{it}^{(l)}$. Consequently, we can use our current information about θ expressed through $\pi(\theta | \mathcal{D})$ and compute the posterior distribution $\pi(P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \theta) | \mathcal{D})$ for each ordinal category D_k . This posterior distribution provides point estimates of these relevant probabilities such as posterior expectations

$$E\left(P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \theta) | \mathcal{D}\right) = \int P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \theta) \pi(\theta | \mathcal{D}) d\theta = P(y_{it} = D_k | \mathbf{x}_{it}^{(l)}, \mathcal{D}), \quad (7)$$

as well as credible intervals for measuring the uncertainty of the estimation.

Our model also allows to explore the estimated relationship between the ordinal and latent variables. We could address the posterior distribution of the latent variable y_{it}^* for each time with regard to each individual in the study or a generic one. The logistic distribution, $\text{Lo}(m_{it}, s = 1)$, for the latent variable y_{it}^* is a conditional distribution with an unknown mean that depends on θ and \mathbf{b}_i . The subsequent marginal distribution $f(y_{it}^* | \mathbf{x}_{it}^{(l)}, \theta)$ can be obtained as in (6) integrating out the random effects and can be interpreted as a time-specific population distribution of the latent variable for a generic individual. Again, this marginal distribution is also a conditional distribution that depends on the population parameters θ and the posterior distribution of the latent variable can be estimated as

$$f(y_{it}^* | \mathbf{x}_{it}^{(l)}, \mathcal{D}) = \int f(y_{it}^* | \mathbf{x}_{it}^{(l)}, \theta) \pi(\theta | \mathcal{D}) d\theta. \quad (8)$$

3.2. Impact of the covariates on the risk of the event

The hazard ratio (HR) of an individual with covariates \mathbf{x} having the event as compared with an individual with covariates \mathbf{x}^* is $\exp\left\{\sum_{p=1}^P \eta_p (x_p - x_p^*)\right\}$, where P is the number of covariates. This hazard ratio only depends on $\boldsymbol{\eta}$, the vector of regression coefficients in (3). Consequently, its posterior distribution,

$$\pi\left(\exp\left\{\sum_{p=1}^P \eta_p (x_p - x_p^*)\right\} | \mathcal{D}\right), \quad (9)$$

computed from the approximate MCMC sample from the posterior marginal $\pi(\boldsymbol{\eta} | \mathcal{D})$, provides all the relevant information about that HR.

The association parameter α allows to assess the relationship of the mean of the latent density with the hazard function but does not provide a direct link with the ordinal longitudinal variable. To facilitate an interpretation of the association parameter in terms of the ordinal measurements, we propose the following *ad-hoc* procedure:

- 1 Compute the posterior mean, $E(\gamma_k | \mathcal{D})$, of the cut-points γ_k and construct the posterior intervals $(E(\gamma_{(k-1)} | \mathcal{D}), E(\gamma_k | \mathcal{D}))$, $k = 1, \dots, K$.
- 2 Define for a given time t a representative value \tilde{m}_{kt} , $k = 1, \dots, K$ of the mean of each ordinal category in the latent scale as follows
 - (a) Compute the median of the posterior distribution (8) in each interval $(E(\gamma_{(k-1)} | \mathcal{D}), E(\gamma_k | \mathcal{D}))$, and consider them, \tilde{y}_{kt}^* , as the representative value of the latent variable y^* in each ordinal category.
 - (b) For each \tilde{y}_{kt}^* , generate a value \tilde{m}_{kt} of the latent mean according to the general formulation (2) $y^* = m + \epsilon$, or equivalently $m = y^* - \epsilon$, where ϵ is a random error with logistic distribution $\text{Lo}(0, 1)$.
- 3 Approximate the conditional HR, given α , of an individual in the ordinal category k having the event versus an individual in category k' at time t as $e^{\alpha(\tilde{m}_{kt} - \tilde{m}_{k't})}$.
- 4 Compute the posterior distribution of the approximate HR in step 3 from the marginal posterior distribution of α as $\int e^{\alpha(\tilde{m}_{kt} - \tilde{m}_{k't})} \pi(\alpha | \mathcal{D}) d\alpha$.

3.3. Prediction

Bayesian reasoning approaches the estimation of the conditional survival probability of an individual i with a given history provided by their baseline covariates and longitudinal follow-up \mathcal{Y}_{in_i} (which guarantees that their survival time is greater than the time, t_{in_i} , of their last longitudinal measurement) through

the posterior distribution $\pi(P(T_i \geq t \mid T_i > t_{in_i}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) \mid \mathcal{D}), t > t_{in_i}$. This posterior contains all relevant information about the location and variability of this conditional survival probability over time. In particular, its posterior mean can be more easily computed as

$$\begin{aligned} P(T_i \geq t \mid T_i > t_{in_i}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \mathcal{D}) &= \int P(T_i \geq t \mid T_i > t_{in_i}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) \pi(\boldsymbol{\theta}, \mathbf{b}_i \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{in_i}) d(\boldsymbol{\theta}, \mathbf{b}_i) \\ &= \int P(T_i \geq t \mid T_i > t_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) \pi(\boldsymbol{\theta}, \mathbf{b}_i \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{in_i}), t > t_{in_i}, \end{aligned} \tag{10}$$

where the conditional probability $P(T_i \geq t \mid T_i > t_{in_i}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i)$ does not depend on the particular longitudinal trajectory, \mathcal{Y}_{in_i} , as a result of the induced independence of the shared effects joint model, and $\pi(\boldsymbol{\theta}, \mathbf{b}_i \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{in_i})$ is the marginal posterior distribution of the common parameters' and random effects' vector for individual i , given \mathcal{Y}_{in_i} and $T_i > t_{in_i}$.

We could also approach prediction of a future longitudinal measurement of an individual in the study [44]. The posterior predictive distribution of a new longitudinal measurement y_{i,n_i+1} at the time t_{i,n_i+1} of a future scheduled appointment for individual i with covariates \mathbf{x}_i and longitudinal ordinal history \mathcal{Y}_{in_i} is given by

$$\begin{aligned} P(y_{i,n_i+1} = D_k \mid T_i > t_{i,n_i+1}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \mathcal{D}) &= P(y_{i,n_i+1}^* \in (\gamma_{k-1}, \gamma_k] \mid T_i > t_{i,n_i+1}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \mathcal{D}) \\ &= \int P(y_{i,n_i+1}^* \in (\gamma_{k-1}, \gamma_k] \mid T_i > t_{i,n_i+1}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) \pi(\boldsymbol{\theta}, \mathbf{b}_i \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{i,n_i+1}) d(\boldsymbol{\theta}, \mathbf{b}_i) \\ &= \int P(y_{i,n_i+1}^* \in (\gamma_{k-1}, \gamma_k] \mid \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) \pi(\boldsymbol{\theta}, \mathbf{b}_i \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{i,n_i+1}) d(\boldsymbol{\theta}, \mathbf{b}_i) \end{aligned} \tag{11}$$

where

$$P(y_{i,n_i+1}^* \in (\gamma_{k-1}, \gamma_k] \mid \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) = \frac{e^{\gamma_k - m_{i,n_i+1}} - e^{\gamma_{k-1} - m_{i,n_i+1}}}{(1 + e^{\gamma_k - m_{i,n_i+1}})(1 + e^{\gamma_{k-1} - m_{i,n_i+1}})}$$

is obtained from (1), with $m_{i,n_i+1} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{i,n_i+1}$. The conditional probability $P(y_{i,n_i+1}^* \in (\gamma_{k-1}, \gamma_k] \mid T_i > t_{i,n_i+1}, \mathcal{Y}_{in_i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i)$ is independent on the survival history, $T_i > t_{i,n_i+1}$, as a consequence of the shared random effects joint model. Note also that the different longitudinal measurements of the same individual are independent given $(\boldsymbol{\theta}, \mathbf{b}_i)$.

The previous two posteriors both apply to individuals in the study and to individuals of the population that could be involved in the study in the future. In this framework, some discussion about the posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{b} \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{in_i})$ becomes necessary. If the interest concentrates on a specific individual in the study, such as individual i , for whom we want to estimate the conditional probability (10) or the predictive distribution (11) from the current data \mathcal{D} , the information provided by \mathcal{Y}_i and $T_i > t_{in_i}$ is already included in \mathcal{D} . Consequently, $\pi(\boldsymbol{\theta}, \mathbf{b} \mid \mathcal{D}, \mathcal{Y}_{in_i}, T_i > t_{in_i}) = \pi(\boldsymbol{\theta}, \mathbf{b} \mid \mathcal{D})$. If the interest focuses on sequentially estimate (10) or/and predict (11) as a result of their follow-up, we would need to sequentially update the current posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{b} \mid \mathcal{D})$ with all that new relevant follow-up information, in particular new longitudinal measurements and the updated survival time.

Dynamic posterior estimation and prediction for individuals of the population who have not participated in the study are also possible. Let us consider now a new subject i' who initially has not participated in the inferential process and enters the study at time $a_{i'}$ with given values $\mathbf{x}_{i'}$ of the baseline covariates. The posterior distribution of their unconditional survival probability is $\pi(P(T_{i'} \geq t \mid T_{i'} > a_{i'}, \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'}) \mid \mathcal{D})$ with posterior mean

$$\begin{aligned} P(T_{i'} \geq t \mid T_{i'} > a_{i'}, \mathbf{x}_{i'}, \mathcal{D}) &= \int P(T_{i'} \geq t \mid T_{i'} > a_{i'}, \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'}) \pi(\boldsymbol{\theta}, \mathbf{b}_{i'} \mid \mathcal{D}, T_{i'} > a_{i'}) d(\boldsymbol{\theta}, \mathbf{b}_{i'}) \\ &= \int \frac{P(T_{i'} \geq t \mid \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'})}{P(T_{i'} > a_{i'} \mid \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'})} \pi(\boldsymbol{\theta}, \mathbf{b}_{i'} \mid \mathcal{D}, T_{i'} > a_{i'}) d(\boldsymbol{\theta}, \mathbf{b}_{i'}), \end{aligned} \tag{12}$$

Prediction of their first longitudinal measurement planned at a fixed time $t_{i'} > a_{i'}$ is

$$\begin{aligned} P(y_{i',1} = D_k | T_{i'} > a_{i'}, \mathbf{x}_{i'}, \mathcal{D}) &= P(y_{i',1}^* \in (\gamma_{k-1}, \gamma_k] | T_{i'} > a_{i'}, \mathbf{x}_{i'}, \mathcal{D}) \\ &= \int P(y_{i',1}^* \in (\gamma_{k-1}, \gamma_k] | T_{i'} > a_{i'}, \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'}) \pi(\boldsymbol{\theta}, \mathbf{b}_{i'} | \mathcal{D}, T_{i'} > a_{i'}) d(\boldsymbol{\theta}, \mathbf{b}_{i'}) \quad (13) \\ &= \int P(y_{i',1}^* \in (\gamma_{k-1}, \gamma_k] | \mathbf{x}_{i'}, \boldsymbol{\theta}, \mathbf{b}_{i'}) \pi(\boldsymbol{\theta}, \mathbf{b}_{i'} | \mathcal{D}, T_{i'} > a_{i'}) d(\boldsymbol{\theta}, \mathbf{b}_{i'}) \end{aligned}$$

If as a consequence of the follow-up of this individual, we would like to compute posterior probabilities of the type (10) and/or (11), the subsequent marginal posterior distribution will come from the joint posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{b}, \mathbf{b}_{i'} | \mathcal{D}, \mathcal{Y}_{i'n_{i'}}, T_{i'} > t_{i'n_{i'}})$, which includes the common parameters and hyper-parameters $\boldsymbol{\theta}$, the vector of random effects \mathbf{b} associated to the original individuals in the study and those of that new individual considered, $\mathbf{b}_{i'}$.

From a Bayesian point of view, the incorporation of sequential information from an individual who is already involved in the study or from the follow-up of a future subject implies the need of sequentially update the posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D})$. In the case of studies based on samples with large sample size, we would expect a minimal change in the estimation of the common parameters but possibly not in the subject specific random effects. The process of updating a posterior distribution for which we only have an approximate random sample and not an analytical distribution is conceptually easy but not so in practice. The main tools to carry out this computational process are based on sequential MCMC methods [45,46] and although are beyond the scope of this paper are a current aim of our research team. Rizopoulos [47] proposes, as an approximation of the subsequent posterior distribution, updating the specific random effects associated with individuals. In particular, the author uses empirical Bayesian estimation for the random effects and an asymptotic normal distribution, based on maximum likelihood estimation, for the common population parameters. Taylor *et al.* [48] also separately update the vector of random effects by using a quick MCMC based on a prior distribution for the population parameters coming from the marginal posterior $\pi(\boldsymbol{\theta} | \mathcal{D})$.

4. Joining longitudinal breast density and age at breast cancer detection

Let $\{D_1, D_2, D_3, D_4\}$ denote the set of BI-RADS breast density categories $\{a, b, c, d\}$, which represent low, medium, high, and very high density, respectively, y_{ij} the breast density category of woman i , $i = 1, \dots, n$, at time t_{ij} (age $50 + t_{ij}$), $j = 1, \dots, n_i$ and y_{ij}^* her subsequent underlying continuous latent value.

Following (3), the connection between both processes is

$$y_{ij} = D_k \Leftrightarrow y_{ij}^* \in (\gamma_{k-1}, \gamma_k], \quad k = 1, 2, 3, 4,$$

where $-\infty = \gamma_0 < \gamma_1 < \gamma_2 < \gamma_3 < \gamma_4 = \infty$ are unknown cut-points, with $\gamma_2 = 0$, and $\text{Lo}(m_{ij}, s = 1)$ represents the corresponding logistic distribution for y_{ij}^* .

Considering the evidence of a decreasing trend of breast density with age and a linear trajectory for the longitudinal latent breast density of woman i

$$(y_i^*(t) | m_{it}) = m_{it} + \epsilon_{it} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + \epsilon_{it}, \quad (14)$$

where $(\beta_0, \beta_1)^T$ and $(b_{i0}, b_{i1})^T$ are the fixed (population) and random effects (individual) for the intercept and the slope term, respectively, and ϵ_{it} the error term. Random effects $(\mathbf{b}_0, \mathbf{b}_1)^T$, where $\mathbf{b}_0 = (b_{01}, \dots, b_{0n})^T$ and $\mathbf{b}_1 = (b_{11}, \dots, b_{1n})^T$, are assumed conditionally i.i.d. with $(b_{i0} | \sigma_0) \sim N(0, \sigma_0)$ and $(b_{i1} | \sigma_1) \sim N(0, \sigma_1)$.

The hazard function of age at breast cancer diagnosis is defined in terms of the left truncated relative risk regression model

$$h_i(t | \mathbf{x}_i, \boldsymbol{\theta}_{is}, t_i^* > a_i) = h_0(t | \lambda, \eta_0) \exp\{\eta_1 \text{Famhist}_i + \eta_2 \text{Brstproc}_i + \alpha m_i(t)\}, \quad t > a_i, \quad (15)$$

where $h_0(t | \lambda, \eta_0) = \lambda t^{\lambda-1} e^{\eta_0}$ is the baseline risk function of a Weibull distribution, $\text{We}(\lambda, e^{\eta_0})$; *family history of breast cancer (Famhist)* and *prior breast procedures (Brstproc)* are dichotomous baseline covariates with associated coefficients η_1 and η_2 , respectively; α assesses the effect of the individual

trajectory of breast density on breast cancer risk in terms of the latent breast density mean; and a_i is the age over 50 at which woman i enters the screening program, thus providing the left truncated time [43].

We assumed prior independence among all the parameters and hyperparameters as a default specification and, with the aim of giving all inferential prominence to the data, we elicited wide proper prior distributions. For the parameters in the longitudinal submodel, we followed Lunn *et al.* [30] except for the standard deviations. In particular, we selected $N(0, 100)$ for the β 's regression coefficients. The ordinal constraint for the cutpoints of the latent scale, $-\infty < \gamma_1 < \gamma_2 = 0 < \gamma_3 < \gamma_4 = \infty$, was expressed by truncating the subsequent prior distributions in the appropriate parametric subspace

$$\begin{aligned} \gamma_1 &\sim N(-\log(3), \sigma_{\gamma_1} = 100) \mathcal{I}(-\infty, \gamma_2 = 0), \\ \gamma_3 &\sim N(\log(3), \sigma_{\gamma_3} = 100) \mathcal{I}(\gamma_2 = 0, \infty), \end{aligned}$$

where $\mathcal{I}(-)$ is the indicator function. Prior means for γ_1 and γ_3 , respectively, correspond to the first and third quartiles of a logistic distribution $\text{Lo}(0, 1)$ in order to provide the same prior probability to each response category. For the standard deviations σ_0 and σ_1 , we choose a uniform distribution, $\text{Un}(0, 10)$. In the case of the survival submodel, we selected $N(0, 100)$ for the η 's regression coefficients as well as for the association coefficient α , and a gamma distribution $\text{Ga}(1, 1)$ for the parameter λ of the baseline hazard function because it mimics a constant baseline hazard function [49].

4.1. Posterior distribution

The posterior distribution $\pi(\theta, \mathbf{b} \mid \mathcal{D})$ was computed in terms of the hierarchical modeling (4) and approximated using MCMC simulation methods through the JAGS software [50]. In particular, we run three MCMC chains with 100,000 iterations, 10,000 of which were used for the burn-in period. The chains were thinned by only storing every 270th iteration in order to reduce autocorrelation in the saved sample. Trace plots of the simulated values of the three chains appear overlapping one another indicating stabilization. Convergence of the chains to the posterior distribution was assessed through the potential scale reduction factor, \hat{R} , and the effective number of independent simulation draws, n_{eff} , [51] and [52], respectively. \hat{R} compares the within-chain variance to the estimated variance of the posterior distribution in such a way that \hat{R} values near 1 indicate that the simulated process has reached the posterior distribution. n_{eff} deals with the level of autocorrelation of the chains simulated values, so that $n_{\text{eff}} > 100$ indicates that sufficient MCMC samples have been obtained.

Table II summarizes the approximate MCMC random sample from the marginal posterior distribution $\pi(\theta \mid \mathcal{D})$ through the mean, median, standard deviation, 2.5% and 97.5% percentiles. The last column of Table II contains the probability that the corresponding parameter is positive: a 0.5 probability would indicate that a positive value of the parameter is equally likely that a negative one, hence indicating little relevance of the corresponding variable (given the remaining covariates). This is not the case for the parameters of our model with probabilities that show a clear *preference* for being above or under zero.

The marginal posterior distribution associated to the population intercept β_0 and slope β_1 of the mean of the latent breast density clearly states that both are negative, $P(\beta_0 < 0 \mid \mathcal{D}) = P(\beta_1 < 0 \mid \mathcal{D}) = 1$, indicating decreasing values over time of the true latent breast density and therefore a higher probability

Table II. Posterior summaries of the parameters and hyperparameters of the breast cancer joint model.

	Mean	SD	2.5%	Median	97.5%	$P(\cdot > 0 \mid \mathcal{D})$
β_0	-1.4262	0.0346	-1.4964	-1.4251	-1.3608	0.0000
β_1	-0.0524	0.0018	-0.0560	-0.0524	-0.0489	0.0000
σ_0	2.6067	0.0227	2.5643	2.6059	2.6534	
σ_1	0.0053	0.0018	0.0015	0.0053	0.0087	
γ_1	-4.6994	0.0269	-4.7521	-4.6998	-4.6489	
γ_3	1.7362	0.0156	1.7060	1.7364	1.7675	
λ	1.5366	0.1044	1.3287	1.5387	1.7386	
η_0	-7.6066	0.3369	-8.2476	-7.6011	-6.9337	0.0000
η_1	0.6227	0.1716	0.2747	0.6308	0.9517	0.9984
η_2	0.4535	0.1440	0.1644	0.4600	0.7210	1.0000
α	0.1490	0.0207	0.1089	0.1496	0.1887	1.0000

of being in the lower breast density categories with age. The variability associated with the random intercept is important, $E(\sigma_0 | \mathcal{D}) = 2.6067$, as a sign of high population heterogeneity with regard to initial breast density. In contrast, there is small variability in the subject-specific slopes, $E(\sigma_1 | \mathcal{D}) = 0.0053$, which denotes that subject-specific trajectories of the true latent breast density do not differ much from the population trend. The estimation of the cut-points γ_1 and γ_3 is very stable and accurate. The posterior means of the coefficients associated to the baseline covariates, *family history of breast cancer*, and *prior breast procedures*, 0.6227 and 0.4535, respectively, indicate an increase of risk of breast cancer detection when women have one or both of these risk factors. These values are consistent with the ones reported in the literature. The strength of the association between the breast density and age at breast cancer diagnosis is assessed through their posterior expectation $E(\alpha | \mathcal{D})=0.149$ and 95% credible interval (0.1089, 0.1887). In addition, the posterior probability 1 for that coefficient being positive provides strong support on the connection between breast density and breast cancer risk.

4.2. Probabilities associated to breast density BI-RADS categories

Figure 2 (top) shows the posterior mean and 95% credible interval of the posterior distribution $\pi(P(y_{it} = D_k | \theta) | \mathcal{D})$ associated to each BI-RADS category for a generic woman in the study. Probabilities associated to category BI-RADS *b* are always higher than 0.5, and grow slightly with age. Probabilities for categories *a*, *c*, and *d* are initially very similar, but categories *c* and *d* decrease with age following a similar pattern while category *a* increases (see Table S1 in Appendix). The information provided by credible intervals is very valuable, thus indicating high precision in the estimated means.

Figure 2 (bottom) shows a violin plot (a combination of a kernel density plot and a boxplot) of the posterior marginal distribution of the latent breast density at ages 50, 55, 60, 65 and 70. The four categories of the ordinal breast density are marked with regard to the posterior mean of the cut points γ_1 and γ_3 , and $\gamma_2 = 0$. The visual comparison between real and latent results is very interesting. We clearly appreciate that the posterior marginal distribution of the latent breast density tends towards lower values with age. In

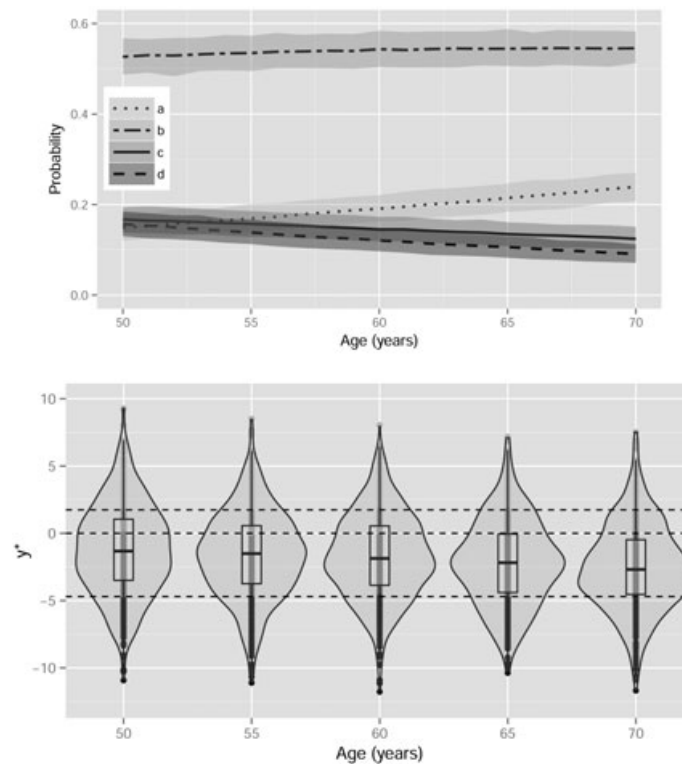


Figure 2. Age-specific population distribution of breast density. Posterior mean and 95% credible interval of the probability associated to each BI-RADS category with respect to age (top) and violin plot of the posterior marginal distribution of the latent breast density of an average woman at ages 50, 55, 60, 65, and 70 (bottom). Horizontal dotted lines represent the posterior mean of the cut-points, thus approximately indicating the region of the latent density corresponding to ordinal BD categories *a*, *b*, *c*, and *d* (from bottom to top).

addition, the bottom tail of the distributions increase with age in detriment of the top tail, thus indicating the general decreasing of breast density with age.

4.3. Assessment of the impact of the study variables on breast cancer risk

Relevant HRs arise from the combination of baseline covariate categories. Figure 3 shows the posterior distribution of the HRs of a breast cancer diagnosis for *family history of breast cancer*, *prior breast procedures*, and both risk factors, with posterior means 1.864, 1.574, and 2.934, respectively. The marginal effects of each covariate are relevant, with posterior probabilities 0.998 and 1.000 associated to HR values greater than 1 for *family history of breast cancer* and *prior breast procedures*, respectively.

Following the *ad-hoc* procedure presented in Section 3.2, Figure 4 shows the posterior mean and 95% credible intervals with regard to age of the approximate HR of a breast cancer diagnosis for women with breast densities *b*, *c*, and *d* compared with women with the same covariate values and breast density *a*. Changes in breast density from category *a* towards more dense categories have a strong effect on breast cancer risk. We observe posterior means of the HR around four for category *d* versus *a*, and HRs greater than 1 (around 1.7 and 2.6) when comparing categories *b* and *c* versus *a*, respectively. A gently wavy behavior for the posterior distributions and credible intervals of the HRs with respect to age can be appreciated, as a consequence of the simulation of the logistic error in the procedure.

4.4. Prediction

Figure 5 shows the posterior mean (10) and 95% credible band for four of the women without cancer in Figure 1, at the end of follow-up. As expected, breast cancer-free probabilities are very high and decrease with age. It is worth noting the narrowness of the bands corresponding to women with a higher number of density measurements, possibly because of the precision of the random effects estimates. Women with high breast density values seem to have lower breast cancer-free survival probabilities. However, we must

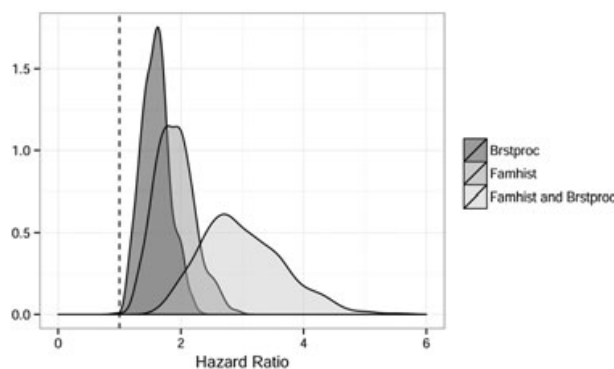


Figure 3. Posterior distribution of the hazard ratios associated to family history of breast cancer and prior breast procedures.

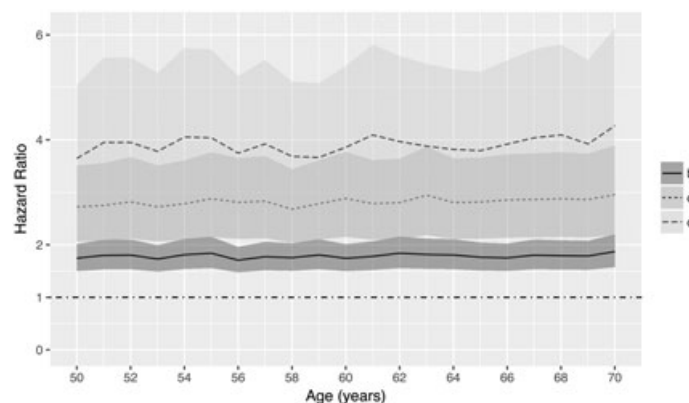


Figure 4. Approximate posterior mean and 95% credible interval with regard to age of the HRs of a BC diagnosis for women with breast density *b*, *c*, and *d* as compared with women with the same covariates and BD measurement *a*.

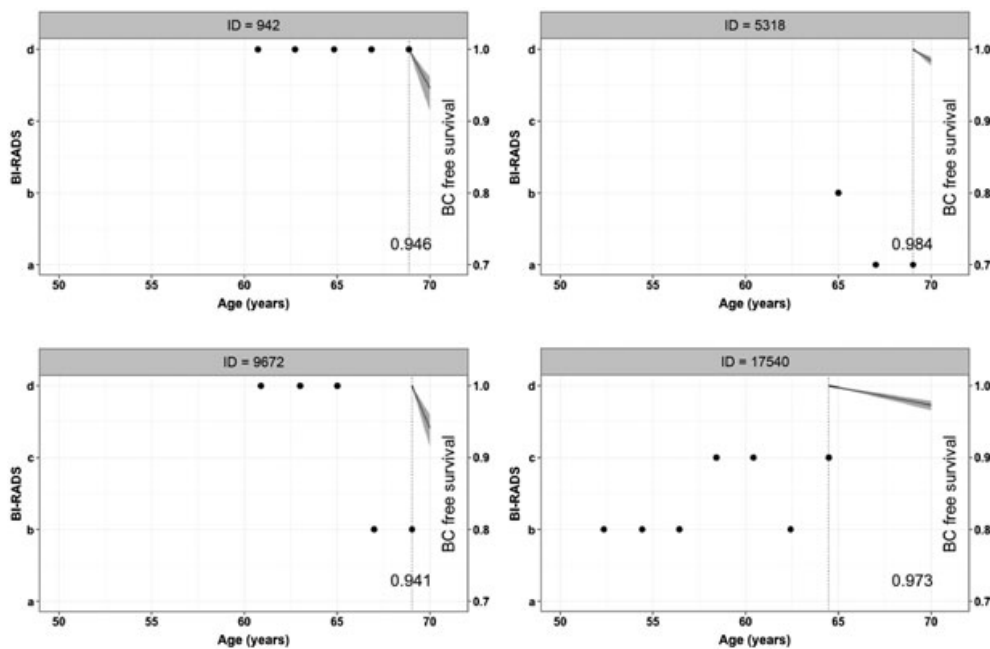


Figure 5. Posterior mean and 95% credible band of the probability of a breast cancer-free diagnosis for women with IDs 942; 5318; 9672; and 17540 without breast cancer at the end of the follow-up. The value of the probability at the lower right of each graphic is the subsequent posterior mean at 70 years.

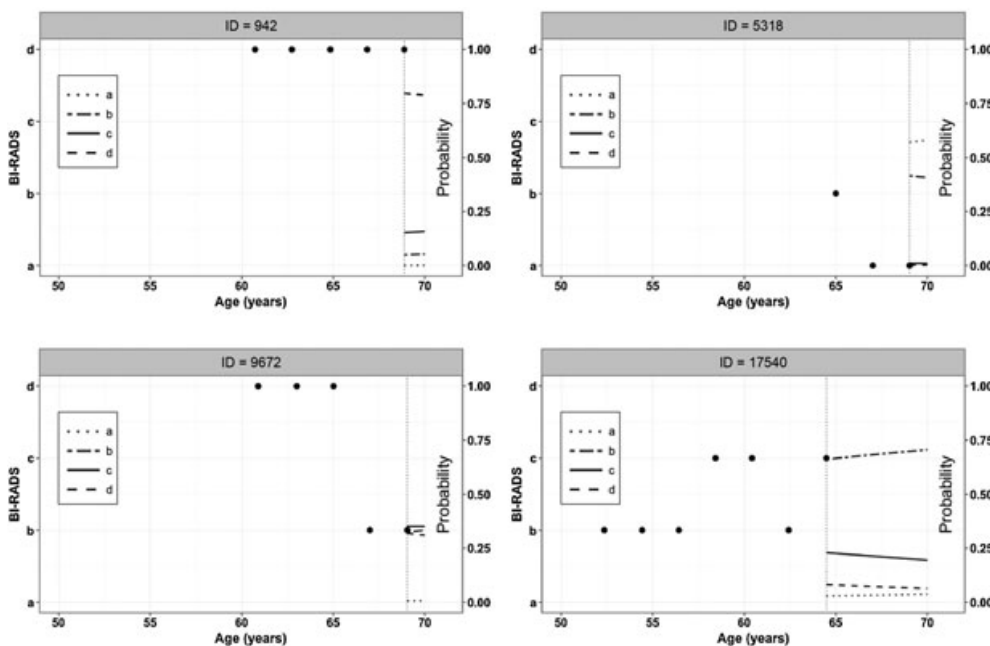


Figure 6. Posterior predicted mean of the breast density in the BI-RADS scale over age for women with IDs 942; 5318; 9672; and 17540 without breast cancer at the end of the follow-up.

also consider the baseline risk factors. Thus, disease-free survival is higher for woman 942, who has a stable very high breast density, than for woman 9672, who experiences a decrease in density. This result can be attributed to presence of prior breast procedures in woman 9672 and absence of them in woman 942. Both women do not have family history of breast cancer.

Figure 6 shows the posterior predictive distribution of the ordinal breast density categories for the women in Figure 5. We appreciate a great variability among the predicted BI-RADS trajectories, and for most of the selected women, category *b* has the highest probabilities over age. But, for woman 5318

category *a* is always the most prevalent with an increasing trend over age, followed by categories *b*, *c* and *d*. These results are in line with the observed breast density trajectory for this woman: three breast density measurements with values *b*, *a* and *a* at late ages 65, 67, and 69, respectively. In contrast, for woman 942 category *d* is always the most prevalent with a decreasing trend over age. These results are also consistent with the observed breast density trajectory for this woman, stable with very high breast density at relatively late ages 61, 63, 65, 67, and 69.

5. Discussion

We propose a Bayesian joint model that combines the information provided by a longitudinal ordinal process and a left-truncated time-to-event outcome. The joint density of both processes is approached through a shared-parameter model which generates a structure of association and conditional independence between both outcomes by means of a vector of common random effects.

We chose a latent variable formulation for the longitudinal process which translated the ordinal scale to the framework of linear mixed models, with a logistic distribution for the measurement error. The latent variable approach facilitates the computational implementation of the model but introduces complexity in the interpretation of results. We assume a logistic distribution for the latent variable that implies the logistic link for cumulative probabilities. Other models might be also appropriate. The most usual alternatives are the normal and the extreme value distributions, which result in the probit and complementary log-log regression links, respectively. It is widely accepted that probit and logit links produce similar results. This also occurs in our study (results not shown), where we have implemented the probit link to assess the robustness of the model. This is not the case for the extreme value distribution that, unlike the logistic and normal ones, is not symmetrical.

We consider that the cut-points that relate the ordinal and latent variables are common for all individuals and time. This assumption may produce some stiffness in the longitudinal model. Thus, individual or time-specific cut-points might endow of more flexibility to the longitudinal latent variable at the expense of a more complex model. Dealing with more than four categories in the ordinal variable is not straightforward. One of the reasons for this is that one or both of the endpoints of the truncated intervals in which the marginal prior distribution of each unknown cut-point is defined can be also unknown [53]. The estimation of these models involves important computational issues in the MCMC sampling that have provided many discussion and proposals, such as hybrid Metropolis–Hasting algorithms to sample from the subsequent posterior distribution [54,55].

Robustness is a major statistical concern in Hierarchical Bayesian models because it can be affected by an inappropriate choice of hyperprior distributions for hyperparameters. We have tested the sensitivity of the model using other prior specifications for the hyperprior distribution of the random effects scale parameter. In particular, we have considered a wide uniform distribution, $Un(0, 100)$, as an alternative to the elicited $Un(0,10)$ in the paper and inverse-gamma hyperdistributions, $IGa(0.01, 0.01)$ and $IGa(0.001, 0.001)$, because of their common use in Bayesian applications. All of them provided almost identical results (not shown in the paper), possibly because of the large sample size.

Our proposal could be applied to a variety of real problems devoted to analyze time-to-event outcomes with temporal ordinal endogeneous covariates. We explored the role of death prior to breast cancer diagnosis as a competing risk. The cumulative incidences estimated with the Kaplan–Meier method or the competing risks with cause-specific hazards approach are very similar, even at older ages (See Figure A1 in the Supplementary file). Therefore, even though the censoring due to the competing risk “death” was informative, it would hardly affect our estimates. Event times have been modeled in terms of relative risk models with left-truncation as a corrective mechanism for the delayed entry bias. Left truncation is common in observational studies of risk factors, where not all the participants enter the study at time zero. We select the Weibull distribution as baseline risk function because it is a traditional model for survival data with a great flexibility in representing different types of risk. The exploration of more sophisticated baseline risk functions, which include multimodality and heavy tailed distributions, in the area of Bayesian joint modeling is a relevant subject with strong connections with the specification of prior distributions. See [56] for a detailed explanation of piecewise constant hazard models and Gamma processes. In addition, the latent linear mixed model is a flexible model that can accommodate heterogeneous trajectories, from linear to complex functions. This is the case of linear models expressed in terms of spline bases to accommodate non-linear profiles [57].

We have used our joint model for analyzing the relationship between mammographic breast density and breast cancer risk in women attending a public screening program. A linear subject-specific trajectory of

the latent variable is included in a relative risks survival model together with two of the most known breast cancer risk factors, *family history of breast cancer* and *previous breast procedures*. Our joint model for breast cancer and breast density is a good starting point that provides results consistent with the literature [16, 18, 19, 21]. They are the basis for a rationale for extending the model and assessing its adequacy and accuracy. Evaluating the ability of breast density to predict time to breast cancer diagnosis, under our joint model, by means of calibration and discrimination measures [58] is a major concern of our research. The discriminative power of our model should offset its complexity. Discrimination measures based on receiver operating characteristic curves are commonly used for assessing predictive accuracy. We are currently exploring a general latent variable approach that could be appropriate.

In contrast to studies published to date, our study is the first to have used the complete longitudinal history of breast density for assessing breast cancer risk over time, at population and individual level. Potential benefits of the proposed joint model include obtaining individual predictions of time-free of breast cancer at age $u > t$, given the observed responses up to age t , and also individual longitudinal predictions of future breast density values. Thus, a joint model similar to that shown here could be used for surveillance of breast cancer risk over time, for scheduling screening exams based on individual dynamic predictions, and also in discussing prevention strategies for those at high risk [16].

Acknowledgements

This paper was partially supported by the research grants Combination and Propagation of Uncertainties (ComPro_UN, MTM2013-42323-P), Statistical methods for clinical trials, complex censoring schemes and integrative omics data analysis (MTM2015-64465-C2-1-R), and Women participation in decisions and strategies on early detection of breast cancer (PI14/00113) from the Spanish Ministry of Economy and Competitiveness, ACOMP/2015/202 from the Generalitat Valenciana, and GRBIO-2014-SGR464 and GRAES-2014-SGR978 from the Generalitat de Catalunya.

We thank Dimitris Rizopoulos and Arindom Chakraborty for their advice on joint modeling. We are also indebted to the research group Grup de Recerca en Anàlisi Estadística de la Supervivència (GRASS), specially to Carles Serrat, for their support and fruitful discussions. We also thank Núria Torà and Marina Pont for their work on data management, and JP Glutting for review and editing. We wish to acknowledge three anonymous referees and the Associate Editor for their useful comments that improved very much the original version of the paper.

We also would like to thank the staff of the Cancer Screening Office and the radiologists of the BCEDP-VOE, for their valuable help in obtaining the data we analyzed. And, last but not least, we appreciate all women that kindly answered the questionnaire.

References

1. Onega T, Beaber EF, Sprague BL, Barlow WE, Haas JS, Tosteson AN, Schnall MD, Armstrong K, Schapira MM, Geller B, Weaver DL, Conant EF. Breast cancer screening in an era of personalized regimens: a conceptual model and National Cancer Institute initiative for risk-based and preference-based approaches at a population level. *Cancer* 2014; **19**: 2955–2964.
2. Vilaprinyo E, Forné C, Carles M, Sala M, Pla R, Castells X, Domingo L, Rue M, Interval Cancer Study Group (INCA). Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PLoS One* 2014; **9**:e86858.
3. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* 1989; **81**:1879–1886.
4. Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. *Statistics in Medicine* 2011; **30**(10):1090–1104.
5. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine* 2004; **23**(7):1111–1130.
6. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinian A. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Research and Treatment* 2012; **133**(1):1–10.
7. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer prediction in a screening population. *Breast Cancer Research and Treatment* 2005; **94**:115–122.
8. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute* 2006; **98**:1215–1226.
9. Barlow WE, White E, Ballard-Barbash R, Vacek P M, Titus-Ernstoff L, Carney PA, Tice JA, Buist DS, Geller BM, Rosenberg R, Yankaskas BC, Kerlikowske K. Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute* 2006; **98**:1204–1214.

10. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Annals of Internal Medicine* 2008; **148**:337–347.
11. Byrne C, Schairer C, Wolfe JN, Parekh N, Salane M, Brinton LA, Hoover RN, Haile R. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *Journal of the National Cancer Institute* 1995; **87**:1622–1629.
12. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 1998; **7**:1133–44.
13. Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiology, Biomarkers & Prevention* 2004; **13**:715–722.
14. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, Pankratz VS. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Research* 2007; **9**:217–225.
15. van Gils CH, Hendriks JHCL, Holland R, Karssemeijer N, Otten JDM, Straatman H, Verbeek ALM. Changes in mammographic breast density and concomitant changes in breast cancer risk. *European Journal of Cancer Prevention* 1999; **8**:509–515.
16. Kerlikowske K, Ichikawa L, Miglioretti DL, Buist DS, Vacek PM, Smith-Bindman R, Yankaskas B, Carney PA, Ballard-Barbash R, Consortium NIOHBCS. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. *Journal of the National Cancer Institute* 2007; **99**:386–395.
17. Vachon CM, Pankratz VS, Scott CG, Maloney SD, Ghosh K, Brandt KR, Milanese T, Carston MJ, Sellers TA. Longitudinal trends in mammographic percent density and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 2007; **16**:921–928.
18. Lokate M, Stellato RK, Veldhuis WB, Peeters PHM, van Gils CH. Age-related changes in mammographic density and breast cancer risk. *American Journal of Epidemiology* 2013; **178**:101–109.
19. Work ME, Reimers LL, Quante AS, Crew KD, Whiffen A, Beth Terry M. Changes in mammographic density over time in breast cancer cases and women at high risk for breast cancer. *International Journal of Cancer* 2014; **135**:1740–1744.
20. Huo CW, Chew GL, Britt KL, Ingman WV, Henderson MA, Hopper JL, Thompson EW. Mammographic density - a review on the current understanding of its association with breast cancer. *Breast Cancer Research and Treatment* 2014; **144**:479–502.
21. Kerlikowske K, Gard CC, Sprague BL, Tice JA, Miglioretti DL. One versus two breast density measures to predict 5- and 10-year breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 2015; **24**(6):889–897.
22. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data* 2nd ed. Wiley: Hoboken, 2002.
23. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journals - American Statistical Association* 1995; **90**:27–37.
24. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**:1663–1685.
25. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
26. Rizopoulos D. *Joint Models For Longitudinal and Time-to-Event Data with Applications in R*. CRC Press, Biostatistics Series: Boca Raton, FL, 2012.
27. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 2010; **35**:1–33.
28. Rizopoulos D. *JMbayes: Joint modeling of longitudinal and time-to-event data under a bayesian approach*, 2015. Available from: <http://CRAN.R-project.org/package=JMbayes>, r package version 0.7-2, [access date December 28, 2015].
29. Serrat C, Rué M, Armero C, Piulachs X, Perpiñán H, Forte A, Páez A, Gómez G. Frequentist and bayesian approaches for a joint model for prostate cancer risk and longitudinal prostate-specific antigen data. *Journal of Applied Statistics* 2015; **42**(6):1223–1239.
30. Lunn DJ, Wakefield J, Racine-Poon A. Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores. *Statistics in Medicine* 2001; **20**:2261–2285.
31. Chakraborty A, Das K. Inferences for joint modelling of repeated ordinal scores and time to event data. *Computational and Mathematical Methods in Medicine* 2010; **11**:281–295.
32. Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine* 2010; **29**:546–557.
33. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine* 2014; **33**:580–594.
34. Hogan JW, Laird N. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **28**:239–257.
35. Chakraborty A. Bounded influence function based inference in joint modelling of ordinal partial linear model and accelerated failure time model. *Statistical Methods in Medical Research* 2014. DOI 0962280214531570.
36. Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statistics in Medicine* 2016; **35**:382–398.
37. He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson disease. *Statistical Methods in Medical Research* 2016; **25**(4):1346–1358.
38. Liu F, Li Q. A Bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. *Statistical Methods in Medical Research* 2016; **25**(5):2180–2192.
39. Baré ML, Montes J, Florensa R, Sentís M, Donoso L. Factors related to non-participation in a population-based breast cancer screening programme. *European Journal of Cancer Prevention* 2003; **12**:487–94.
40. Baré M, Bonfill X, Andreu X. Relationship between the method of detection and prognostic factors for breast cancer in a community with a screening programme. *Journal of Medical Screening* 2006; **13**:183–91.

41. Baré M, Sentís M, Galceran J, Ameijide A, Andreu X, Ganau S, Tortajada L, Planas J. Interval breast cancers in a community screening programme: frequency, radiological classification and prognostic factors. *European Journal of Cancer Prevention* 2008; **17**:414–21.
42. American College of Radiology. Breast Imaging Reporting and Data System®(BI-RADS®) 2013.
43. Uzunoğullari a, Wang J-L. Comparison of hazard rate estimators for left truncated and right censored data. *Biometrika* 1992; **79**:297–310.
44. Armero C, Forte A, Perpiñán H, Sanahuja MJ, Agustí S. Bayesian joint modeling for assessing the progression of Chronic Kidney Disease in children. *Statistical Methods in Medical Research* 2016. DOI:10.1177/0962280216628560.
45. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2006; **68**:411–436.
46. Andrieu C, Doucet A, Holenstein R. Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72**:269–342.
47. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**:819–829.
48. Taylor JMG, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; **69**(1):206–213.
49. Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* 2004; **58**:1–9.
50. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria, 2003.
51. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**:457–511.
52. Kass RE, Carlin BP, Gelman A, Neal R. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician* 1998; **52**:93–100.
53. Albert JH, Siddhartha C. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* 1993; **88**:669–679.
54. Cowles MK. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* 1996; **6**:101–111.
55. Nandram B, Chen MH. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Statistics in Medicine* 1996; **54**:129–144.
56. Ibrahim JG, Chen M-H, Sinha D. *Bayesian Survival Analysis*. Springer: New York, 2001.
57. Andrinopoulou E-R, Rizopoulos D, Takkenberg JJM, Lesaffre E. Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine* 2014; **33**(18):3167–3178.
58. Rizopoulos D, Verbeke G, Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2010; **66**:20–29.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Supplementary material for “Bayesian joint ordinal and survival modeling for breast cancer risk assessment”

C. Armero⁽¹⁾, C. Forné^(2,3), M. Rué^(2,4), A. Forte⁽¹⁾, H. Perpiñán^(1,5), G. Gómez⁽⁶⁾, and M. Baré^(4,7)

In this supplemental document, we provide additional data analysis results and the BUGS code of the proposed model for studying breast cancer diagnosis with regard to longitudinal breast density measurement.

BI-RADS and age

Table A1. Posterior expectation of the probability associated to each BI-RADS category with respect to age.

BI-RADS	Age (years)				
	50	55	60	65	70
<i>a</i>	0.150	0.168	0.191	0.214	0.239
<i>b</i>	0.526	0.538	0.541	0.545	0.545
<i>c</i>	0.166	0.157	0.147	0.136	0.125
<i>d</i>	0.158	0.137	0.121	0.106	0.092

⁽¹⁾ Department of Statistics and Operational Research, Universitat de València, Doctor Moliner, 50, 46100-Burjassot, Spain.

⁽²⁾ Department of Basic Medical Sciences, Universitat de Lleida-IRBLleida, Avda. Rovira Roure, 80, 25198-Lleida, Spain.

⁽³⁾ Oblikue Consulting, Barcelona, Spain.

⁽⁴⁾ Health Services Research Network in Chronic Diseases (REDISSEC).

⁽⁵⁾ Fundación para el Fomento de la Investigación Sanitaria y Biomédica (FISABIO), Generalitat Valenciana, Spain.

⁽⁶⁾ Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain.

⁽⁷⁾ Clinical Epidemiology and Cancer Screening. Corporació Sanitària Parc Taulí-UAB, Sabadell, Parc Taulí s/n, 08208, Sabadell, Spain.

Cumulative incidence

In Figure A1, the dashed line shows the cumulative incidence calculated using the Kaplan-Meier approach without accounting for competing risk events. The solid line shows the cumulative incidence, after adjusting for competing risk. The dotted line shows the cumulative incidence of the death occurring prior to the breast cancer diagnosis.

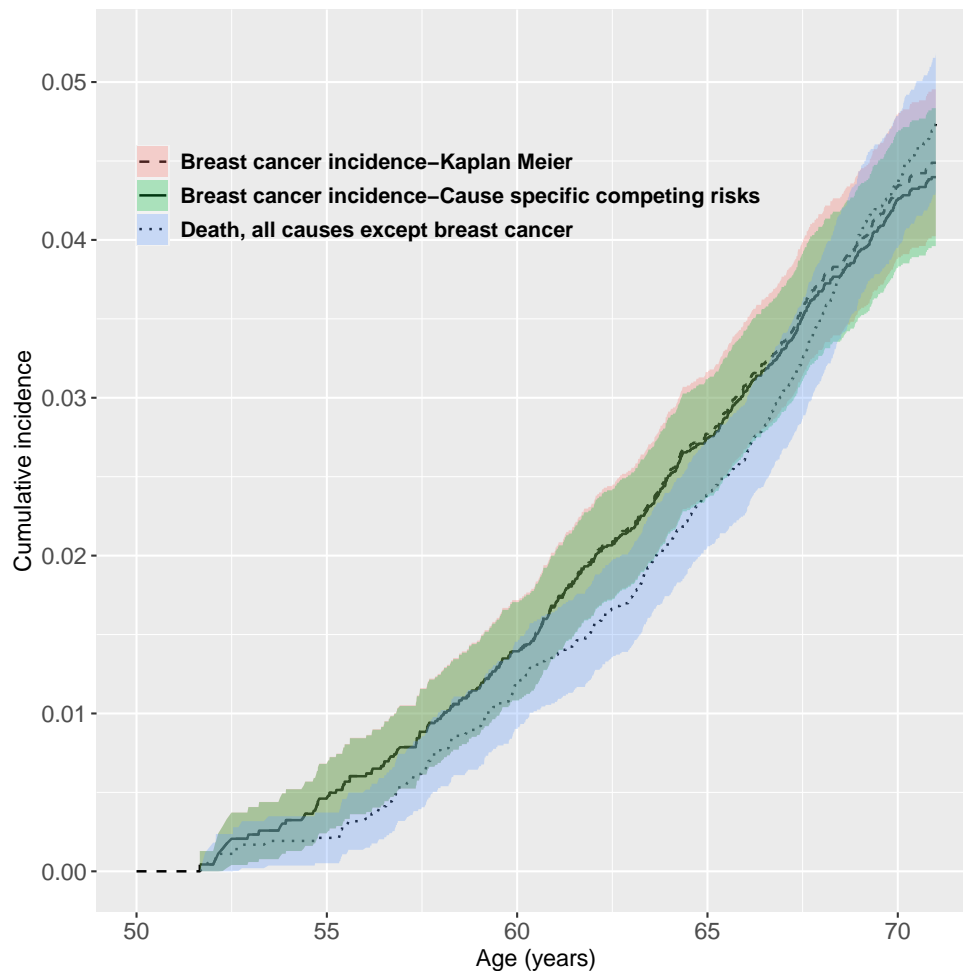


Figure A1. The estimated cumulative incidence of breast cancer using the Kaplan-Meier approach and the competing risks approach.

JAGS code for the proposed model

```

model{
for(i in 1:N){
  for(j in 1:K[i]){
    y[i,j] ~ dcat(p[i,j,1:4]) #1
    p[i,j,1] <- 1-q[i,j,1]
    p[i,j,2] <- q[i,j,1]-q[i,j,2]
    p[i,j,3] <- q[i,j,2]-q[i,j,3]
    p[i,j,4] <- q[i,j,3]
  }
}

```



```

logit(q[i,j,1]) <- location[i,j]-gamma1
logit(q[i,j,2]) <- location[i,j]
logit(q[i,j,3]) <- location[i,j]-gamma3
location[i,j] <- beta0 + b0[i] + beta1*time[i,Proj[i,j]]
                + b1[i]*time[i,Proj[i,j]]
}#times[i,] a grid of times for individual i and times[i,Proj[i,j]]
is the observed time of individual i at time j

#0nes' trick for a not implemented likelihood in BUGS
ones[i] ~ dbern(prob[i]) #2
prob[i] <- L[i]/C

## Likelihood for survival data - event=1 if event, and 0 if censoring
L[i] <- (pow(f[i],event[i])*pow(S[i],1-event[i]))/(S0[i])

## Hazard for individual i at each time point of the grid time[i,]
for(j in 1:J[i]){
  h[i,j] <- (lambda*pow(time[i,j], (lambda-1))) * exp(eta[1] + eta[2]*numrel[i]
              + eta[3]*brstproc[i] + alpha*(beta0 + b0[i] + (beta1 + b1[i])*time[i,j]))
}

## Cumulative hazard H[t] = int_0^t (h[u] du) using GAUSS-LEGENDRE polynomials
and evaluated at 12 different (n.end) points indicated by Proj.aux
for(j in 1:n.end){ #3
  Aux[i,j] <- ((time[i,Proj.aux[j+1]]-time[i,Proj.aux[j]])/2)
              *inprod(wk.int, h[i, (Proj.aux[j]+1):(Proj.aux[j]+5)])
              #wk.int Gauss-Legendre weights
  H[i,j] <- sum(Aux[i,1:j])
}

## Survival probability for individual i at their final time
S[i] <- exp(-H[i,n.end])
## Survival probability for individual i at their entry time
S0[i] <- exp(-H[i,n.ini])

## Density function
f[i] <- h[i,J[i]]*S[i]

b0[i] ~ dnorm(0, tau0)
b1[i] ~ dnorm(0, tau1)
}
C <- 10000

#Prior distributions for the survival process
for(i in 1:3){ eta[i] ~ dnorm(0.0, 0.0001) }

```

```
alpha ~ dnorm(0.0, 0.0001)
lambda ~ dgamma(1, 1)

#Prior distributions for the longitudinal process
beta0 ~ dnorm(0.0, 0.0001)
beta1 ~ dnorm(0.0, 0.0001)
gamma1 ~ dnorm(-1.0986, 0.001)I(,0)
gamma3 ~ dnorm(1.0986, 0.001)I(0,)
tau0 <- pow(sig0, -2)
sig0 ~ dunif(0, 100)
tau1 <- pow(sig1, -2)
sig1 ~ dunif(0, 100)
}
```

1. Specification of the latent variable
2. One's trick for the Survival likelihood to incorporate left truncation.
3. Code for the numerical integration for $H(t)$ using Gauss-Legendre polynomials and evaluated at different time points

3.3 Mètodes i resultats de l'objectiu 3

En aquesta secció es presenten els mètodes utilitzats i els resultats obtinguts per donar resposta al tercer dels objectius plantejats en aquesta tesi. Tota la metodologia i els resultats estan continguts en el manuscrit original Forné *et al.* (Clinical Kidney Journal 2019), la informació del qual es detalla a continuació, junt amb el resum traduït al català.

Títol: Machine learning analysis of serum biomarkers for cardiovascular risk assessment in chronic kidney disease

Autors: Forné C, Cambray S, Bermudez-Lopez M, Fernandez E, Bozic M, Valdivielso JM; on behalf of the NEFRONA investigators

Revista: Clinical Kidney Journal 2019;sfz094

Factor d'impacte de l'any anterior de la publicació: 2,975 (percentil 76,9 a *Urology & Nephrology*)

DOI: 10.1093/ckj/sfz094

Resum:

Antecedents: Els pacients amb malaltia renal crònica (MRC) presenten una major càrrega d'ateroesclerosi i un alt risc d'esdeveniment cardiovascular. S'han descrit diversos biomarcadors associats a esdeveniments cardiovasculars, però la seva efectivitat combinada, en l'estratificació del risc cardiovascular, no ha estat provada encara. L'objectiu del present treball és analitzar la capacitat combinada de 19 biomarcadors associats a la malaltia ateromatosa en la predicció d'esdeveniments cardiovasculars després de 4 anys de seguiment en una subcohort de l'estudi NEFRONA, en individus amb diferents estadis d'MRC sense esdeveniment cardiovascular previ.

Mètodes: Es van quantificar dinou suposats biomarcadors en 1.366 pacients (73 esdeveniments cardiovasculars), i la seva capacitat predictiva d'esdeveniments cardiovasculars es va avaluar mitjançant models de bosc aleatori de supervivència. Els factors associats amb el risc cardiovascular van ser contrastats en models de regressió de Fine i Gray, amb la mort no cardiovascular i el trasplantament renal com a esdeveniments competitiu.

Resultats: L'anàlisi de supervivència per bosc aleatori va identificar diversos biomarcadors com a rellevants per predir esdeveniments cardiovasculars. La inclusió d'aquests biomarcadors en un model Fine i Gray va mostrar un augment del risc cardiovascular per nivells alts d'osteopontina, d'osteoprotegerina, de metalloproteïnasa de matriu 9 i del factor de creixement de l'endoteli vascular, tot i que només va millorar de manera marginal la discriminació obtinguda amb els paràmetres clínics clàssics [índex de concordança (interval de confiança del 95%): 0,744 (0,609, 0,878) i 0,723 (0,592, 0,854), respectivament]. Tanmateix, en individus amb diabetis tractats amb antihipertensius i



hipolipemians, la determinació d'aquests biomarcadors podria ajudar a millorar les estimacions de risc cardiovascular.

Conclusions: La determinació de quatre biomarcadors en sèrum de pacients amb MRC podria millorar la predicció del risc cardiovascular en individus d'alt risc.



ORIGINAL ARTICLE

Machine learning analysis of serum biomarkers for cardiovascular risk assessment in chronic kidney disease

Carles Forné ^{1,2,*}, Serafi Cambay^{3,*}, Marcelino Bermudez-Lopez³, Elvira Fernandez³, Milica Bozic³ and Jose M. Valdivielso ³ on behalf of the NEFRONA investigators**

¹Biostatistics Unit, Institute for Biomedical Research Dr. Pifarré Foundation, IRBLleida, Lleida, Spain,

²Department of Basic Medical Sciences, University of Lleida, Lleida, Spain and ³Vascular and Renal Translational Research Group, Institute for Biomedical Research Dr. Pifarré Foundation, IRBLleida and RedinRen RETIC, ISCIII, Lleida, Spain

Correspondence and offprint requests to: Jose M. Valdivielso; E-mail: valdivielso@irbllleida.cat, Milica Bozic; E-mail: bozicm@medicina.udl.cat

*These authors contributed equally to this work.

**Listed in the Appendix.

ABSTRACT

Background. Chronic kidney disease (CKD) patients show an increased burden of atherosclerosis and high risk of cardiovascular events (CVEs). There are several biomarkers described as being associated with CVEs, but their combined effectiveness in cardiovascular risk stratification in CKD has not been tested. The objective of this work is to analyse the combined ability of 19 biomarkers associated with atheromatous disease in predicting CVEs after 4 years of follow-up in a subcohort of the NEFRONA study in individuals with different stages of CKD without previous CVEs.

Methods. Nineteen putative biomarkers were quantified in 1366 patients (73 CVEs) and their ability to predict CVEs was ranked by random survival forest (RSF) analysis. The factors associated with CVEs were tested in Fine and Gray (FG) regression models, with non-cardiovascular death and kidney transplant as competing events.

Results. RSF analysis detected several biomarkers as relevant for predicting CVEs. Inclusion of those biomarkers in an FG model showed that high levels of osteopontin, osteoprotegerin, matrix metalloproteinase-9 and vascular endothelial growth factor increased the risk for CVEs, but only marginally improved the discrimination obtained with classical clinical parameters: concordance index 0.744 (95% confidence interval 0.609–0.878) versus 0.723 (0.592–0.854), respectively. However, in individuals with diabetes treated with antihypertensives and lipid-lowering drugs, the determination of these biomarkers could help to improve cardiovascular risk estimates.

Conclusions. We conclude that the determination of four biomarkers in the serum of CKD patients could improve cardiovascular risk prediction in high-risk individuals.

Keywords: biomarkers, cardiovascular risk, cohort study, competing risks, random forest

Received: 26.2.2019; Editorial decision: 24.6.2019

© The Author(s) 2019. Published by Oxford University Press on behalf of ERA-EDTA.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

INTRODUCTION

Chronic kidney disease (CKD) is a highly prevalent disease worldwide with a median prevalence of 7.2% in people >30 years of age, but increasing to 23–36% in individuals >64 years of age [1]. These numbers are expected to increase dramatically in the next decades, mainly due to the aging of the population and to the increase of other risk factors such as diabetes, hypertension and obesity [2]. These trends raise significant concern due to the fact that CKD patients have increased risk of major cardiovascular events (CVEs), as cardiovascular disease (CVD) is the leading cause of death in this population [3]. The increased cardiovascular risk in CKD patients seems to be associated, at least in part, with their propensity to present atheromatous plaques, as the prevalence of atherosclerotic-related events is very high [4] and the presence of plaques predicts the incidence of CVEs [5]. Today, prevention, diagnosis and prognosis of CVD in CKD patients is difficult, as the standard risk scores underestimate the cardiovascular risk in CKD patients [6, 7].

With the aim of modifying or improving current cardiovascular risk assessment algorithms, there has been an extensive search for serum biomarkers useful to predict CVEs associated with atheromatosis [8]. Atherosclerotic plaque is a manifestation of a systemic disease. Consequently, associated biomarkers can be the product of the disease itself or derive directly from the plaque [9]. Thus, published studies have focused on markers of plaque presence, instability, inflammation, neovascularization, calcification, stenosis severity and atheromatous CVEs [10, 11]. However, one of the main problems related with the search for new biomarkers is that they are usually evaluated in an isolated context and only corrected by the confounding effect of traditional risk factors in multivariable regression analysis. To date, no study has evaluated the possible role of a combination of a large number of biomarkers to predict CVEs in a CKD population.

Machine learning algorithms are powerful tools to make predictions based on large amounts of data. Among them, random forest algorithms build several decision trees. Each tree is built based on a random subsample and the rest of the sample (out-of-bag) is used to assess the model performance and the variable importance. For each node, a subset of variables is explored to determine which variable maximizes the improvement of the outcome prediction. Finally, by ensembling the results of all trees, measures of performance can be obtained [concordance index (C-index), variable importance, etc.]. Therefore, random forest could be a useful tool to assess the contribution of several biomarkers on atheromatous disease-related events.

In this context, this study was carried out with the main objective of evaluating whether 19 biomarkers previously associated with atherosclerotic disease can increase the capacity of traditional risk factors for predicting CVEs in patients with different stages of CKD and controls included in the NEFRONA cohort.

MATERIALS AND METHODS

Study design

This study included 1366 subjects from the NEFRONA [12, 13] study of which 813 presented with CKD (380 Stage 3, 304 Stages 4 and 5 and 129 dialysis) and 553 were controls without CKD (estimated glomerular filtration rate based on creatinine >60 mL/min/1.73 m²). Briefly, NEFRONA is a multi-centre, observational, prospective study designed to assess the predictive value of

non-invasive imaging techniques and biomarkers for CVEs and mortality in CKD patients. Patients 18–75 years of age and at different CKD stages were recruited in Spain from 2009 to 2012. Exclusion criteria were a history of CVEs, significant carotid stenosis, active infections (human immunodeficiency virus and tuberculosis), pregnancy, life expectancy <12 months and having received any organ transplantation or carotid artery surgery. A 4-year follow-up was performed collecting data on CVEs, both fatal and non-fatal. CVEs were defined according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Physicians responsible for the patients' recruitment recorded the CVEs. In case of out-of-hospital death, family members were interviewed to determine death circumstances. Each local ethics committee approved the study and all included patients signed an informed consent.

Clinical data and biochemical variables

Anthropometric data [sex, age and body mass index (BMI)], clinical data [systolic and diastolic blood pressure (SBP and DBP), pulse pressure (PP)], family history of early CVD, cardiovascular risk factors (smoking, diabetes, hypertension and dyslipidaemia) and blood samples were collected by a specifically trained itinerant team formed by two technicians and a specifically trained nurse. The diagnoses of diabetes, hypertension and dyslipidaemia were obtained directly from the clinical records. Biochemical parameters were obtained from routine blood tests 3 months before or after an ultrasound was performed.

Collected blood samples were stored at –80°C in the Biobank of the RedInRen in the University of Alcalá de Henares (Madrid) until biomarker determination, which was performed in a blinded manner. All biomarkers were measured using MILLIPLEX MAP kits (EMD Millipore, Burlington, MA, USA). The list of biomarkers assessed, the minimum detectable level of each biomarker as well as the information about intra- and interassay coefficients of variation are shown in the [Supplementary data, Table S1](#).

We selected biomarkers associated with processes of atherosclerosis, focusing on inflammatory chemokines [14], proteins related to bone metabolism and calcification [15] and proteins related to plaque inflammation, progression and vascularization [10]. The other needed criteria used were the availability of MILLIPLEX MAP kits for quantification.

Atherosclerosis assessment

Atherosclerosis assessment was performed at baseline as previously described [16]. Participants underwent a carotid and femoral ultrasound in three territories of both carotid arteries (internal and common carotid arteries and carotid bulbs) and two territories of both femoral arteries (common and superficial femoral arteries). Plaques were defined according to the American Society of Echocardiography and the Mannheim carotid intima-media thickness (cIMT) Consensus as cIMT lumen protrusion ≥1.5 mm [17, 18]. Plaque presence was defined as plaque in at least one territory. Detailed protocols for plaque imaging, quantification and prevalence in the NEFRONA cohort have been published previously [19].

Statistical analysis

Qualitative variables were described with frequencies and percentages and quantitative variables with medians and 25th and 75th percentiles. Bivariate analysis by occurrence of CVEs was performed by means of the logrank test.

To discover new biomarkers for cardiovascular incidence, we performed random survival forest (RSF) analysis for competing risks [20]. The competing events to the CVEs were non-cardiovascular death and kidney transplant. We fitted two models: one with known risk factors and confounders—without biomarkers—and the other adding new potential biomarkers. Previously the same algorithm was used for missing values imputation. Additionally, quantitative variables were previously centred and scaled. No variables were excluded from analysis by near-zero variance or high pairwise correlation among predictors. Forests were grown using the modified Gray's splitting rule, selecting variables based on their direct effect on the cumulative cardiovascular incidence. Variables were ranked by the mean decrease in prediction error [measure of variable importance (VIMP)] and the C-index) was obtained as a measure of model performance. Confidence intervals (CIs) for VIMP were calculated subsampling the survival forest [21]. Finally, we included in a Fine and Gray (FG) competing risk regression model the variables with the 5th percentile of VIMP > 0, thus ensuring 95% confidence of being relevant to the outcome of interest. The final model was selected by backward stepwise regression based on the minimum Akaike information criterion (AIC). We evaluated the calibration and discrimination of the final models. Moreover, model assumptions were assessed (linearity and proportional subhazards). When needed for validating the models, transformations of variables were used. We also provide a nomogram of the model predictions in order to simplify its interpretation.

All tests were two-sided at a significance level $\alpha = 0.05$. All statistical analyses were carried out using R statistical software (R Foundation for Statistical Computing, Vienna, Austria). The randomForestSRC package was used for RSF algorithm implementation, the crrstep and riskRegression packages for FG modelling and the crskdiag package for model assumptions validation.

RESULTS

Bivariate analysis to identify biomarkers of CVEs

Over a median follow-up time of 48 months, 73 CVEs were reported: 13 in the control group, 22 in CKD Stage 3, 27 in CKD Stages 4 and 5 and 11 in the dialysis group. Higher hazards of CVEs were found for older patients, for patients with plaque at baseline and for patients in more advanced CKD stage. Moreover, patients with or being treated for diabetes, hypertension and dyslipidaemia or with the presence of atrial fibrillation showed higher hazards of having a CVE. Patients with an increased risk of a CVE also showed low levels of high-density lipoprotein (HDL) cholesterol and 25-hydroxyvitamin D3 [25(OH)D3], higher levels of triglycerides, high-sensitivity C-reactive protein (hsCRP), phosphate, potassium, SBP and PP (Table 1). In addition, cardiovascular risk was also higher for patients with higher levels of soluble vascular cell adhesion molecule (sVCAM-1), matrix metalloproteinase-9 (MMP-9), osteopontin (OPN), osteocalcin (OC), vascular endothelial growth factor (VEGF), osteoprotegerin (OPG) and monocyte chemoattractant protein 1 (MCP-1) (Table 2).

RSF analysis to identify biomarkers of CVEs

RSF performed without biomarkers showed that the 10 most important variables for CVE prediction were SBP, PP, diabetes,

Table 1. Characteristics of participants according to the occurrence of CVEs

Variable	No event (n = 1293)	Event (n = 73)	P-value
Age (years)	59 (49, 67)	63 (56, 70)	<0.001
Sex (female), n (%)	557 (43.1)	27 (37.0)	0.286
BMI (kg/m ²)	27.9 (25, 31.3)	29.6 (25.8, 32.4)	0.065
Dyslipidaemia, n (%)	695 (53.8)	55 (75.3)	<0.001
Lipid-lowering drugs, n (%)	642 (49.7)	55 (75.3)	<0.001
Hypertension, n (%)	871 (67.4)	66 (90.4)	<0.001
Antihypertensive drugs, n (%)	850 (65.7)	64 (87.7)	<0.001
Diabetes, n (%)	264 (20.4)	32 (43.8)	<0.001
Antidiabetic drugs, n (%)	235 (18.2)	30 (41.1)	<0.001
Atrial fibrillation, n (%)	16 (1.24)	3 (4.11)	0.030
Heart failure, n (%)	17 (1.31)	1 (1.37)	0.943
Family history of early CVD, n (%)	147 (11.4)	11 (15.1)	0.303
Smoking status, n (%)			0.136
Non-smoker	545 (42.2)	26 (35.6)	
Current smoker	498 (38.5)	26 (35.6)	
Former smoker	250 (19.3)	21 (28.8)	
CKD stage, n (%)			<0.001
Control	540 (41.8)	13 (17.8)	
CKD Stage 3	358 (27.7)	22 (30.1)	
CKD Stages 4 and 5	277 (21.4)	27 (37)	
Dialysis	118 (9.1)	11 (15.1)	
Presence of basal plaque, n (%)	811 (62.7)	60 (82.2)	0.001
Total cholesterol (mg/dL)	190 (163, 214)	180 (156, 209)	0.101
Missing values	46	1	
HDL cholesterol (mg/dL)	49 (41, 61)	44 (34, 52)	<0.001
Missing values	158	12	
LDL cholesterol (mg/dL)	112 (90, 136)	111 (75, 133)	0.168
Missing values	179	14	
Triglycerides (mg/dL)	112 (81, 162)	138 (94, 181)	0.008
Missing values	72	4	
SBP (mmHg)	136 (124, 151)	149 (134, 168)	<0.001
DBP (mmHg)	80 (73, 87)	85 (76, 89)	0.071
PP (mmHg)	55 (46, 67)	68 (53, 79)	<0.001
Potassium (mEq/L)	4.65 (4.31, 5)	4.8 (4.49, 5.3)	0.001
Missing values	88	2	
Phosphate (mg/dL)	3.7 (3.2, 4.2)	3.85 (3.4, 4.4)	<0.001
Missing values	250	5	
Calcium (mg/dL)	9.4 (9.1, 9.7)	9.4 (9, 9.8)	0.851
Missing values	203	5	
hsCRP (mg/L)	1.93 (0.95, 4.13)	2.25 (1.02, 6.68)	0.004
Missing values	15	3	
25(OH)D3 (ng/L)	17.3 (12.7, 21.9)	14.7 (9.5, 20.2)	0.003
Missing values	11	2	

Italic numbers are frequencies of missing data.

Values are shown as medians and 25th and 75th percentiles unless stated otherwise. P-values correspond to the logrank test.

LDL, low-density lipoprotein.

being treated for dyslipidaemia, hypertension, DBP, being treated for diabetes, CKD stage, being treated for hypertension and HDL cholesterol (Figure 1a). When biomarkers were included alongside clinical variables in the RSF analysis, OPN was the most relevant predictor of the incidence of CVEs, followed by PP, SBP and diabetes. OPG also appeared as a potential predictor of cardiovascular risk (Figure 1b). VIMP values for all variables are shown in the Supplementary data, Table S2. Overall discrimination without biomarkers was 0.678 (95% CI 0.649–0.710) and improved when biomarkers were included [0.716 (95% CI 0.689–0.746)].

Table 2. Biomarker levels according to the occurrence of CVEs

Variable	No event (n = 1293)	Event (n = 73)	P-value
Eotaxin (pg/mL)	122 (84, 170)	122 (77, 168)	0.780
Missing values	54	4	
FGF-2 (pg/mL)	60 (33, 100)	65 (39, 114)	0.456
Missing values	50	4	
Fractalkine (pg/mL)	83 (48, 133)	77 (51, 124)	0.929
Missing values	53	4	
GM-CSF (pg/mL)	8.6 (4.2, 16.5)	8.5 (2.7, 16.2)	0.538
Missing values	53	4	
IFN- γ (pg/mL)	5.2 (2.9, 10.6)	5.4 (1.9, 12.3)	0.557
Missing values	53	4	
IP-10 (pg/mL)	446 (309, 689)	451 (323, 704)	0.271
Missing values	53	4	
Leptin (pg/mL)	15 627 (6822, 33 667)	16 601 (8523, 38 389)	0.754
Missing values	37	3	
MCP-1 (pg/mL)	357 (232, 534)	358 (256, 535)	0.028
Missing values	54	4	
MDC (pg/mL)	941 (709, 1220)	903 (721, 1107)	0.463
Missing values	54	4	
MIP-1 β (pg/mL)	32.4 (20.6, 47.9)	30.2 (21.3, 47.1)	0.685
Missing values	53	4	
MMP-9 (pg/mL)	71 (46, 1354)	798 (58, 2071)	0.001
Missing values	38	3	
MPO (ng/mL)	27.1 (0.37, 55.8)	4.6 (0.30, 37.3)	0.746
Missing values	43	3	
OC (pg/mL)	11 501 (8226, 17 022)	14 660 (8915, 22 649)	<0.001
Missing values	37	4	
OPG (pg/mL)	473 (320, 601)	538 (370, 699)	<0.001
Missing values	37	3	
OPN (pg/mL)	13 209 (6049, 23 539)	24 530 (11 471, 43 840)	<0.001
Missing values	38	3	
PAI-1 (pg/mL)	80 (29, 521)	262 (36, 773)	0.074
Missing values	37	3	
sICAM-1 (pg/mL)	147 (107, 624)	293 (108, 758)	0.243
Missing values	38	3	
sVCAM-1 (pg/mL)	27 (17, 8889)	7449 (20, 10 747)	0.002
Missing values	37	3	
VEGF (pg/mL)	81 (38, 153)	88 (45, 161)	0.046
Missing values	52	4	

Italic numbers are frequencies of missing data.

Values are shown as medians and 25th and 75th percentiles. P-values correspond to the logrank test.

FGF-2, fibroblast growth factor 2; GM-CSF, granulocyte-macrophage colony-stimulating factor; IP-10, IFN- γ inducible protein 10; MDC, macrophage-derived chemokine; MIP-1 β , macrophage inflammatory protein-1beta; MPO, myeloperoxidase; PAI-1, plasminogen activator inhibitor-1; sICAM-1, soluble intercellular adhesion molecule-1.

With the most promising variables (those with 95% of VIMP values >0), we performed a multivariate FG competing risks model for cardiovascular incidence (Table 3). Of the traditional risk factors considered, being treated with lipid-lowering drugs and having lower levels of HDL cholesterol showed a significantly increased risk for CVEs. Regarding biomarkers, higher values of OPG, MMP-9, VEGF and OPN were statistically significant predictors of increased cardiovascular risk. Other predictors [interferon gamma (IFN- γ), diabetes, being treated for hypertension, total cholesterol and age] were kept in the model as their inclusion improved the AIC. In order to meet the model assumptions, base-2 logarithm transformations for OPG and OPN were used.

The discrimination ability of the competing risks model without biomarkers at 48 months was moderately high [0.723

(95% CI 0.592–0.854)]. As shown in Figure 2, the inclusion of biomarkers slightly increased the C-index [0.744 (95% CI 0.609–0.878)]. Figure 3 shows the adjusted cumulative cardiovascular incidence, according to the FG model with biomarkers, corresponding to four risk groups defined by the presence or absence of diabetes and being or not treated with both lipid-lowering and antihypertensive drugs. Significant biomarkers—OPN, OPG, MMP-9 and VEGF—were set to their 25th and 75th percentiles. The patients with diabetes and treated with both drugs are those suffering major impacts on absolute cardiovascular risk of higher levels of biomarkers.

Finally, to better visualize our model predictions, we obtained a nomogram illustrating a representative patient of our cohort (Figure 4). The cumulative incidence estimate of CVEs at 48 months for this representative patient was 0.041.

DISCUSSION

In this study we explored the role of 19 serum biomarkers previously shown to be associated with atheromatosis in improving CVE prediction in a cohort mainly composed of CKD patients. The main result of our analysis is that the addition of biomarkers marginally improved the discrimination ability obtained with only traditional risk factors in CKD. Despite this, OPN was the most relevant predictor of CVEs in RSF analysis. Thus when biomarkers were not included in RSF, clinical parameters with higher prediction ability were PP and SBP, data that agree with previous results [22], and the third most relevant clinical parameter was diabetes, also linked to CVEs in the CKD population [23]. Moreover, when the most promising variables in the RSF analysis were included in an FG regression, higher levels of OPN, OPG, VEGF and MMP-9 were significantly associated with a higher incidence of CVEs, along with traditional factors such as HDL cholesterol and being treated for dyslipidaemia.

Our group has previously identified predictors of CVEs in the NEFRONA cohort. First, we focused on clinical variables and routine clinical analytics to find predictors of CVEs. Sex, CKD, 25(OH)-vitamin D₃, cholesterol and subclinical atheromatosis were identified as the main predictors [5]. The inclusion of biomarkers in this analysis overruled the effect of all of them, maybe reflecting that the biomarkers selected have a strong relationship with the atherosclerotic process. Moreover, we cannot rule out that the inclusion of medications also reduced the effect of some clinical variables. In a second work [11] we identified OPN and OPG as factors significantly associated with CVEs. soluble tumor necrosis factor-like weak inducer of apoptosis (sTWEAK) also showed an independent association, but was not included in this work as it was missed in 50% of our samples. In the mentioned study, since we only tried to identify whether they were independently associated with the events, we did not test the improvement of prediction added by the biomarkers or rank the variables as we did in the present work. Furthermore, in this study we identified that patients with diabetes and treated for dyslipidaemia and hypertension could benefit from the determination of biomarkers, as these patients are suffering major impacts on absolute cardiovascular risk of higher biomarker levels. It should be noted that being treated for dyslipidaemia or hypertension can be considered similar to presenting with these comorbidities.

The search for serum biomarkers with potential prognostic ability for CVEs is a very active field of study. However, although results sometimes are confirmed in independent cohorts, a very small number make it to the clinic [24]. This fact could be due to

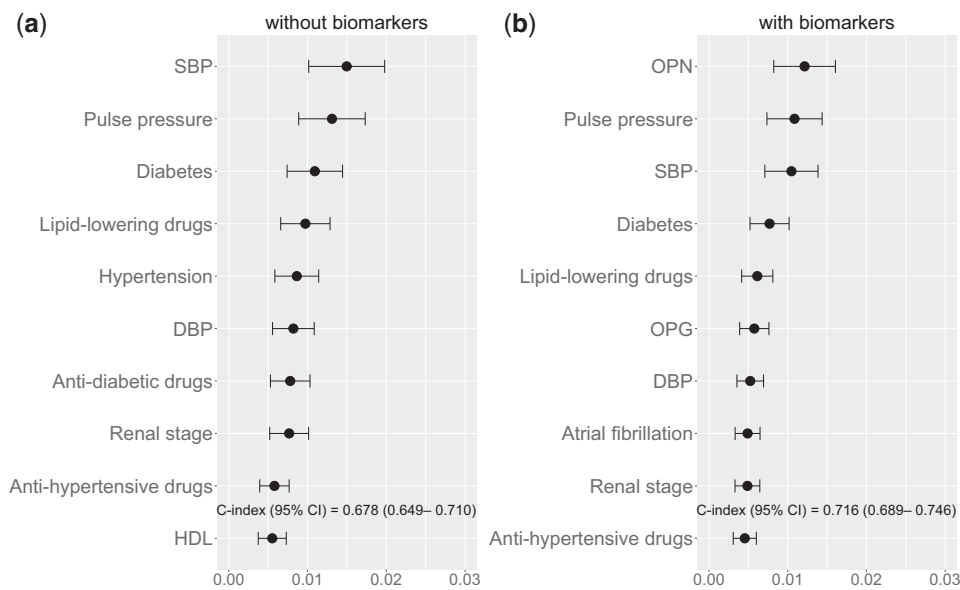


FIGURE 1: Variable importance from the RSF analysis for cardiovascular risk: the top 10 features ranked by mean decrease in prediction error. (a) RSF including only clinical variables and (b) the top 10 features considering clinical variables and biomarkers.

Table 3. Multivariate FG competing risks model for cardiovascular incidence

Variable	Hazard ratio ^a (95% CI)	P-value
log ₂ OPG	12.3 (3.39–44.6)	<0.001
MMP-9	1.24 (1.07–1.42)	0.003
VEGF	1.27 (1.06–1.53)	0.010
log ₂ OPN	1.33 (1.07–1.66)	0.011
HDL cholesterol	0.69 (0.50–0.96)	0.026
Lipid-lowering drugs	1.90 (1.07–3.35)	0.028
SBP	1.25 (0.99–1.58)	0.059
IFN- γ	0.86 (0.73–1.01)	0.072
Diabetes	1.65 (0.95–2.88)	0.075
Antihypertensive drugs	1.74 (0.88–3.44)	0.110
Total cholesterol	1.19 (0.93–1.53)	0.160
Age	1.20 (0.90–1.60)	0.200

^aHazard ratio corresponds to an increase of 1 standard deviation (SD) for continuous predictors (except OPG and OPN). For OPG and OPN, hazard ratio corresponds to an increase of 2-fold difference in their scaled values. Variable (SD): MMP-9 (1360), VEGF (361), HDL cholesterol (15.5), SBP (20.3), IFN- γ (159), total cholesterol (38.8), age (12.2), OPG (414), OPN (28 657).

the very high level of heterogeneity of the populations and to the existence of many confounding factors that are not taken into account in the data collection [25]. Furthermore, many biomarkers might be related between them due to biological interactions in the signalling pathways that cannot be identified when analysed in an independent manner.

By applying RSF, we found a marginal improvement in accuracy when adding potential biomarkers to the traditional risk factors. However, our data show that OPN has high prediction ability for CVE incidence, even higher than well-described clinical variables such as SBP and diabetes [3, 26, 27]. OPN is an extracellular matrix glycoprotein that participates in cell-to-cell contact mechanisms; its levels increase in different models of atherosclerosis [28] and it has been found to be associated with cardiovascular risk in CKD patients [29]. The other biomarker

with good performance in prognosis, when considered together with clinical variables, is OPG. The glycoprotein OPG is a member of the tumour necrosis factor-related family, and part of the receptor activator of nuclear factor-kappaB (RANK)–RANK ligand (RANKL)–OPG network. Higher levels of OPG have been associated with coronary atherosclerosis in several studies [30, 31] and in CKD, dialysis and kidney transplant patients it has been further associated with adverse events [32–34]. These associations were also found with the FG analysis, so both OPN and OPG were predictors included in the final model for cardiovascular risk.

The final FG competing risk model also included MMP-9 and VEGF. MMP-9 has been associated with plaque instability and CVEs [35], and its serum levels have been found to be correlated to carotid atherosclerosis in CKD patients [36]. VEGF is an angiogenic molecule crucial for endothelial integrity and it also contributes to plaque formation and destabilization [26], its levels are increased in CKD [27] and it has been identified in haemorrhagic arterial plaques of CKD patients [37]. We also plotted a nomogram to visualize our model predictions, allowing us to quickly predict the cumulative risk of CVEs at 48 months for any patient by knowing only the 12 selected variables.

One strength of this study is the large sample size, including participants without previous CVEs of a prospective observational study, and the consideration of 19 different biomarkers together with clinical variables. Thus, including the most promising features in cardiovascular risk assessment could improve the prediction ability of current risk scales. Random forest is a versatile analysis method that usually provides good results when compared with alternative algorithms, such as support vector machines, neural networks or some boosting algorithm, which also usually take much longer to build and tune. The random forest also has the advantage of offering greater interpretability. Thus we believe that the RSF algorithm used in this study is useful and appropriate. We would like to point out that the mean decrease in prediction error was chosen as a measure of variable importance in order to rank variables in RSF since

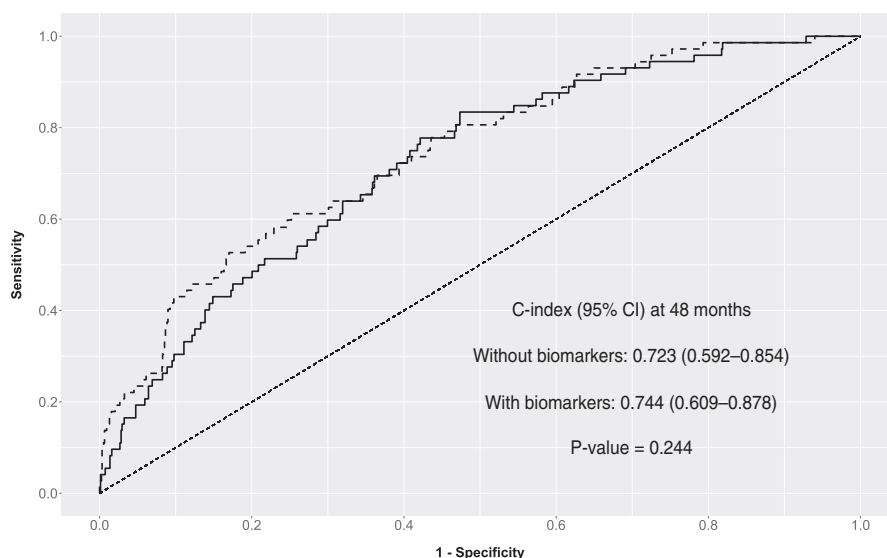


FIGURE 2: ROC curve for prognostic values from both FG regression models, without and with biomarkers.

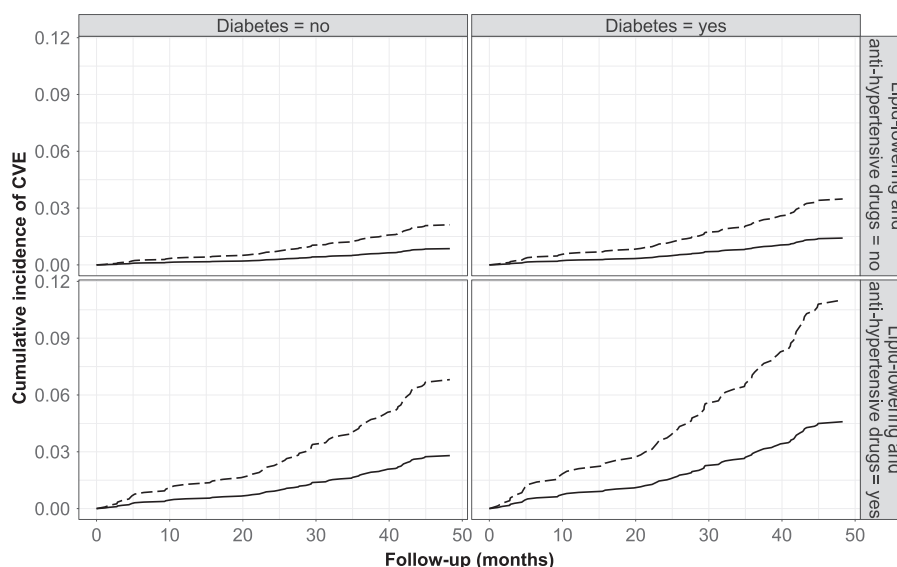


FIGURE 3: Adjusted cumulative cardiovascular incidence obtained with the FG regression model with biomarkers (Table 3). Population was stratified according to diabetes and being treated with lipid-lowering and antihypertensive drugs. Quantitative variables (except significant biomarkers) were set at their median value: HDL cholesterol = -0.104 , SBP = -0.119 , IFN- γ = -0.084 , total cholesterol = -0.024 , age = 0.178 . Significant biomarkers at their 25th percentile (solid line) and 75th percentile (dashed line): OPG = -0.471 and 0.195 , MMP-9 = -0.618 and 0.371 , VEGF = -0.323 and 0.039 , OPN = -0.552 and 0.095 .

that is the most common measure used for this purpose. Nevertheless, other measures have been proposed in RSF analysis [38], so different rankings could have been obtained if other measures had been used. Also, there are alternatives to selecting the FG regression model. Other possible criteria could have resulted in models with different predictors. Finally, we want to point out that the prediction ability of the clinical factors and the biomarkers included in this study could vary with different follow-up times.

Our study has several limitations. First, the small number of CVEs may have limited the capacity to more clearly identify predictors of cardiovascular risk and therefore limit the conclusions of this study. Machine learning algorithms, like random forest used in this analysis, help to partly overcome this limitation and allow identification of strong predictors in problems

with bad signal:noise ratios. Second, detection of the biomarkers was performed by multiplex analysis. Although the method is very useful to detect multiple analytes in small amounts of sample, it should be validated with individual enzyme-linked immunosorbent assay (ELISA) kits. A third limitation is that, as a consequence of the NEFRONA study design, subjects that received a kidney transplant during the follow-up were not followed for CVE assessment, but this limitation was overcome by performing cumulative incidence analysis with a competing risks approach. Furthermore, only a portion of the cohort had samples available to perform the multiplex analysis, so the subsample could not be representative of the whole cohort. Another important limitation is the lack of data on coronary calcium scores, which have shown great power in CVE prediction in the CKD population [39]. Other variables such as

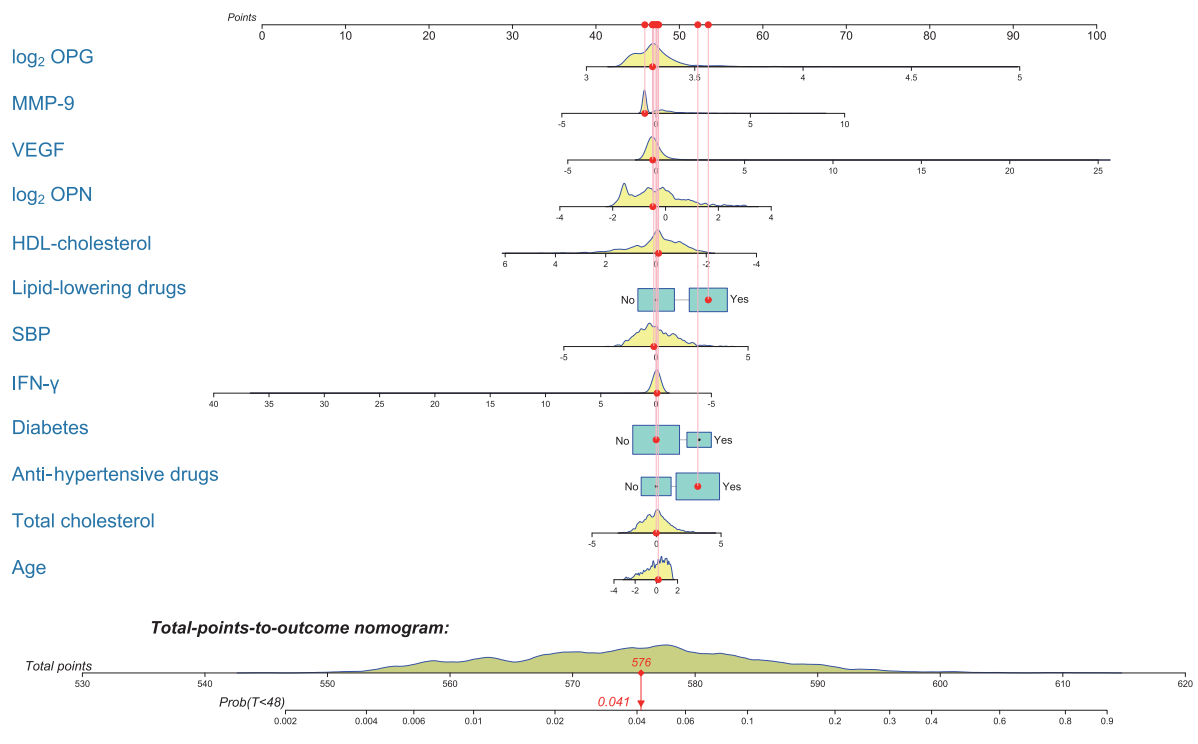


FIGURE 4: Nomogram for predicting cumulative risk of CVDs at 48 months with the FG model. As a cohort's representative, we illustrated a patient with the cohort median values for continuous variables, without diabetes and being treated with lipid-lowering and antihypertensive drugs. With these characteristics, the probability of suffering a CVD after 4 years is estimated to be 4.1%.

changes in medication or poor BP control may affect the outcome and were not recorded. In addition, a single measurement of biomarkers may not be representative and may not directly predict the outcome. Finally, we would like to mention that out-of-hospital death causes were based on family interviews, which could introduce some bias in our study.

In summary, our study confirmed previous results obtained about clinical risk factors with the capacity to predict CVDs in CKD patients. Although a deep and flexible analysis by RSF found no relevant improvement in discrimination ability added by the potential biomarkers assessed in this study, OPN was identified as the best biomarker for possible use in future scales or algorithms to predict CVDs in CKD patients. The parametric competing risks modelling determined that, among all the biomarkers studied, OPN, OPG, VEGF and MMP-9 could be used in combination to predict CVDs in CKD patients. Individuals with diabetes, dyslipidaemia and hypertension could benefit from interventions lowering levels of these biomarkers to decrease cardiovascular risk.

SUPPLEMENTARY DATA

Supplementary data are available at [ndt online](http://ndt.oxfordjournals.org/).

FUNDING

This study was funded by the intramural programme of the IRBLleida, the Instituto de Salud Carlos III (RETIC RD16/0009, PI15/00960, and PI16/01354) and FEDER funds.

AUTHORS' CONTRIBUTIONS

J.M.V., M.B., S.C. and C.F. contributed to the study concept and design. M.B. was involved in data acquisition. J.M.V., S.C., E.F. and M.B.-L. were responsible for data interpretation. C.F., S.C. and J.M.V. contributed to statistical analysis. S.C., C.F. and J.M.V. contributed to drafting of the manuscript. All authors contributed to critical revision of the manuscript for important intellectual content. J.M.V. contributed to study supervision. NEFRONA investigators collected baseline and prospective follow-up data of the cohort.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Zhang QL, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health* 2008; 8: 117
- Webster AC, Nagler EV, Morton RL et al. Chronic kidney disease. *Lancet* 2016; 389: 1238–1252
- Sarnak MJ, Levey AS, Schoolwerth AC et al. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. *Circulation* 2003; 108: 2154–2169
- Thompson S, James M, Wiebe N et al. Cause of death in patients with reduced kidney function. *J Am Soc Nephrol* 2015; 26: 2504–2511

5. Valdivielso JM, Betriu A, Martinez-Alonso M et al. Factors predicting cardiovascular events in chronic kidney disease patients. Role of subclinical atheromatosis extent assessed by vascular ultrasound. *PLoS One* 2017; 12: e0186665
6. Weiner DE, Tighiouart H, Elsayed EF et al. The Framingham predictive instrument in chronic kidney disease. *J Am Coll Cardiol* 2007; 50: 217–224
7. Herzog CA, Asinger RW, Berger AK et al. Cardiovascular disease in chronic kidney disease. A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int* 2011; 80: 572–586
8. Hofer IE, Steffens S, Ala-Korpela M et al. Novel methodologies for biomarker discovery in atherosclerosis. *Eur Heart J* 2015; 36: 2635–2642
9. Tomez MI, Narula J, Kovacic JC. Advances in the understanding of plaque composition and treatment options: year in review. *J Am Coll Cardiol* 2014; 63: 1604–1616
10. Ammirati E, Moroni F, Norata GD et al. Markers of inflammation associated with plaque progression and instability in patients with carotid atherosclerosis. *Mediators Inflamm* 2015; 2015: 718329
11. Bozic M, Méndez-Barbero N, Gutiérrez-Muñoz C et al. Combination of biomarkers of vascular calcification and sTWEAK to predict cardiovascular events in chronic kidney disease. *Atherosclerosis* 2018; 270: 13–20
12. Junyent M, Martínez M, Borrás M et al. [Usefulness of imaging techniques and novel biomarkers in the prediction of cardiovascular risk in patients with chronic kidney disease in Spain: the NEFRONA project]. *Nefrología* 2010; 30: 119–126
13. Junyent M, Martínez M, Borrás M et al. Predicting cardiovascular disease morbidity and mortality in chronic kidney disease in Spain. The rationale and design of NEFRONA: a prospective, multicenter, observational cohort study. *BMC Nephrol* 2010; 11: 14
14. Wan W, Murphy PM. Regulation of atherogenesis by chemokines and chemokine receptors. *Arch Immunol Ther Exp (Warsz)* 2013; 61: 1–14
15. Jono S, Shioi A, Ikari Y et al. Vascular calcification in chronic kidney disease. *J Bone Miner Metab* 2006; 24: 176–181
16. Gracia M, Betriu A, Martínez-Alonso M et al. Predictors of subclinical atheromatosis progression over 2 years in patients with different stages of CKD. *Clin J Am Soc Nephrol* 2016; 11: 287–296
17. Stein JH, Korcarz CE, Hurst RT et al. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force Endorsed by the Society for Vascular Medicine. *J Am Soc Echocardiogr* 2008; 21: 93–111
18. Touboul PJ, Hennerici MG, Meairs S et al. Mannheim carotid intima-media thickness and plaque consensus (2004–2006–2011). An update on behalf of the advisory board of the 3rd, 4th and 5th watching the risk symposia, at the 13th, 15th and 20th European Stroke Conferences, Mannheim, Germany, 2004, Brussels, Belgium, 2006, and Hamburg, Germany, 2011. *Cerebrovasc Dis* 2012; 34: 290–296
19. Betriu A, Martínez-Alonso M, Arcidiacono MV et al. Prevalence of subclinical atheromatosis and associated risk factors in chronic kidney disease: the NEFRONA study. *Nephrol Dial Transplant* 2014; 29: 1415–1422
20. Ishwaran H, Gerds TA, Kogalur UB et al. Random survival forests for competing risks. *Biostatistics* 2014; 15: 757–773
21. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* 2018; 38: 558–582
22. Minutolo R, Gabbai FB, Agarwal R et al. Assessment of achieved clinic and ambulatory blood pressure recordings and outcomes during treatment in hypertensive patients with CKD: a multicenter prospective cohort study. *Am J Kidney Dis* 2014; 64: 744–752
23. Yuan J, Zou X-R, Han S-P et al. Prevalence and risk factors for cardiovascular disease among chronic kidney disease patients: results from the Chinese cohort study of chronic kidney disease (C-STRIDE). *BMC Nephrol* 2017; 18: 23
24. Everett BM, Ridker PM. Biomarkers for cardiovascular screening: progress or passé? *Clin Chem* 2017; 63: 248–251
25. Drucker E, Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J* 2013; 4: 7
26. Holm PW, Slart RH, Zeebregts CJ et al. Atherosclerotic plaque development and instability: a dual role for VEGF. *Ann Med* 2009; 41: 257–264
27. Anderson CE, Hamm LL, Batuman G, et al. The association of angiogenic factors and chronic kidney disease. *BMC Nephrol* 2018; 19: 117
28. Waller AH, Sanchez-Ross M, Kaluski E et al. Osteopontin in cardiovascular disease: a potential therapeutic target. *Cardiol Rev* 2010; 18: 125–131
29. Barreto DV, Lenglet A, Liabeuf S et al. Prognostic implication of plasma osteopontin levels in patients with chronic kidney disease. *Nephron Clin Pract* 2011; 117: c363–c372
30. Jono S, Ikari Y, Shioi A et al. Serum osteoprotegerin levels are associated with the presence and severity of coronary artery disease. *Circulation* 2002; 106: 1192–1194
31. Schoppet M, Sattler AM, Schaefer JR et al. Increased osteoprotegerin serum levels in men with coronary artery disease. *J Clin Endocrinol Metab* 2003; 88: 1024–1028
32. Yilmaz MI, Siritopol D, Saglam M et al. Osteoprotegerin in chronic kidney disease: associations with vascular damage and cardiovascular events. *Calcif Tissue Int* 2016; 99: 121–130
33. Svensson M, Dahle DO, Mjoen G et al. Osteoprotegerin as a predictor of renal and cardiovascular outcomes in renal transplant recipients: follow-up data from the ALERT study. *Nephrol Dial Transplant* 2012; 27: 2571–2575
34. Scialla JJ, Kao WHL, Crainiceanu C et al. Biomarkers of vascular calcification and mortality in patients with ESRD. *Clin J Am Soc Nephrol* 2014; 9: 745–755
35. Yabluchanskiy A, Ma Y, Iyer RP et al. Matrix metalloproteinase-9: many shades of function in cardiovascular disease. *Physiology (Bethesda)* 2013; 28: 391–403
36. Addabbo F, Mallamaci F, Leonardis D et al. Searching for biomarker patterns characterizing carotid atherosclerotic burden in patients with reduced renal function. *Nephrol Dial Transplant* 2007; 22: 3521–3526
37. Nakano T, Ninomiya T, Sumiyoshi S et al. Chronic kidney disease is associated with neovascularization and intraplaque hemorrhage in coronary atherosclerosis in elders: results from the Hisayama Study. *Kidney Int* 2013; 84: 373–380
38. Ishwaran H, Kogalur U, Gorodeski E et al. High-dimensional variable selection for survival data. *J Am Stat Assoc* 2010; 105: 205–217
39. Chen J, Budoff MJ, Reilly MP et al. Coronary artery calcification and risk of cardiovascular disease and death among patients with chronic kidney disease. *JAMA Cardiol* 2017; 2: 635–643

APPENDIX

NEFRONA investigators

The authors thank Eva Castro, Virtudes María, Teresa Molí, Teresa Vidal, Meritxell Soria and the Biobank of RedInRen for their invaluable support. The NEFRONA study investigator group is composed of the following: José Aladrén Regidor (Hospital Comarcal Ernest Lluch, Calatayud); Jaume Almirall, Esther Ponz (Corporació Parc Taulí, Barcelona); Jesús Arteaga Coloma (Hospital de Navarra, Pamplona); Auxiliadora Bajo Rubio, Raquel Díaz (Hospital La Paz, Madrid); Montserrat Belart Rodríguez (Sistemes Renals, Lleida); Antonio Gascón (Hospital Obispo Polanco, Teruel); Jordi Bover Sanjuan (Fundació Puigvert, IIB Sant Pau, Barcelona); Josep Bronsoms Artero (Clínica Girona, Girona); Juan B. Cabezero Romero, Salomé Muray Cases (Hospital Reina Sofía, Murcia); Jesús Calviño Varela (Hospital Universitario Lugus Augusti, Lugo); Pilar Caro Acevedo (Clínica Ruber, Madrid); Jordi Carreras Bassa (Diaverum Baix Llobregat, Barcelona); Aleix Cases Amenós, Elisabet Massó Jiménez (Hospital Clínic, Barcelona); Rosario Moreno López (Hospital de la Defensa, Zaragoza); Secundino Cigarrán Guldris, Saray López Prieto (Hospital Da Costa, Lugo); Lourdes Comas Mongay (Hospital General de Vic, Barcelona); Isabel Comerma (Hospital General de Manresa, Barcelona); Teresa Compte Jové (Hospital Santa Creu Jesús, Tarragona); Marta Cuberes Izquierdo (Hospital Reina Sofía, Navarra); Fernando de Álvaro, Covadonga Hevia Ojanguren (Hospital Infanta Sofía, Madrid); Gabriel de Arriba de la Fuente (Hospital Universitario Guadalajara, Guadalajara); Dolores del Pino y Pino (Complejo Hospitalario Universitario Torrecardenas, Almería); Rafael Diaz-Tejeiro Izquierdo, Francisco Ahijado Hormigos (Hospital Virgen de la Salud, Toledo); Marta Dotori (USP Marbella, Málaga); Verónica Duarte (Hospital de Terrassa, Barcelona); Sara Estupiñan Torres (Hospital Universitario Canarias, Santa Cruz de Tenerife); José Fernández Reyes (Hospital de Segovia, Segovia); Loreto Fernández Rodríguez (Hospital Príncipe de Asturias, Madrid); Guillermina Fernández (Clínica Santa Isabel, Sevilla); Antonio Galán Serrano (Hospital General Universitario de Valencia, Valencia); Cesar García Cantón (Hospital Universitario Insular de Gran Canaria, Las Palmas); Antonio L. García Herrera (Hospital Universitario Puerto Real, Cádiz); Mercedes García Mena (Hospital San Juan de Dios, Zaragoza); Luis Gil Sacaluga, María Aguilar (Hospital Virgen del Rocío, Sevilla); José Luis Górriz (Hospital Universitario Doctor Peset, Valencia); Emma Huarte Loza (Hospital San Pedro, Logroño); José Luis Lerma (Hospital Universitario Salamanca, Salamanca); Antonio Liebana Cañada (Hospital de Jaén, Jaén); Jesús Pedro Marín Álvarez (Hospital San Pedro de Alcántara, Cáceres); Nádía Martín Alemany (Hospital Josep Trueta, Girona); Jesús Martín

García (Hospital Nuestra Señora de Sonsoles, Ávila); Alberto Martínez Castela (Hospital Universitari de Bellvitge, Barcelona); María Martínez Villaescusa (Complejo Hospitalario Universitario de Albacete, Albacete); Isabel Martínez (Hospital Galdakao, Bilbao); Iñigo Moina Eguren (Hospital Basurto, Bilbao); Silvia Moreno Los Huertos (Hospital Santa Bárbara, Soria); Ricardo Mouzo Mirco (Hospital El Bierzo, Ponferrada, León); Antonia Munar Vila (Hospital Universitari Son Espases, Palma de Mallorca); Ana Beatriz Muñoz Díaz (Hospital Virgen del Consuelo, Valencia); Juan F. Navarro González (Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife); Javier Nieto, Agustín Carreño (Hospital General Universitario de Ciudad Real, Ciudad Real); Enrique Noiva Fernández (Complejo Hospitalario de Ourense, Ourense); Alberto Ortiz, Beatriz Fernandez (IIS-Fundación Jiménez Díaz, Madrid); Vicente Paraíso (Hospital Universitario del Henares, Madrid); Miguel Pérez Fontán (Complejo Hospitalario Universitario A Coruña, Coruña); Ana Peris Domingo (Hospital Francisc de Borja, Valencia); Celestino Piñera Haces (Hospital Universitario Marqués de Valdecilla, Santander); Dolores Prados Garrido (Hospital Universitario San Cecilio, Granada); Mario Prieto Velasco (Hospital de León, León); Carmina Puig Marí (Hospital d'Igualada, Barcelona); Maite Rivera Gorrín (Hospital Universitario Ramón y Cajal, Madrid); Esther Rubio (Hospital Puerta del Hierro, Madrid); Pilar Ruiz (Hospital Sant Joan Despí Moisès Broggi, Barcelona); Mercedes Salgueira Lazo, Ana Isabel Martínez Puerto (Hospital Virgen Macarena, Sevilla); José Antonio Sánchez Tomero (Hospital Universitario de la Princesa, Madrid); José Emilio Sánchez (Hospital Universitario Central de Asturias, Oviedo); Ramon Sans Lorman (Hospital de Figueras, Girona); Ramon Saracho (Hospital de Santiago, Vitoria); María Sarrias, Daniel Serón (Hospital Universitari Vall d'Hebron, Barcelona); María José Soler, Clara Barrios (Hospital del Mar, Barcelona); Fernando Sousa (Hospital Rio Carrión, Palencia); Daniel Toran (Hospital General de Jerez, Cadiz); Fernando Tornero Molina (Hospital de Sureste, Arganda del Rey); José Javier Usón Carrasco (Hospital Virgen de la Luz, Cuenca); Ildfonso Valera Cortes (Hospital Virgen de la Victoria, Málaga); Merce Vilaprinyo del Perugia (Institut Catala d'Urologia i Nefrologia, Barcelona); Rafael C. Virto Ruiz (Hospital San Jorge, Huesca); Vicente Pallarés Carratalá (Clínica MEDEFIS, Vila-real, Castellón); Carlos Santos Altozano (CS Azuqueca de Henares, Guadalajara); Miguel Artigao Ródenas (CS Zona III, Albacete); Inés Gil Gil (Área Básica Sanitaria de Arán, CAP Viella, Lleida); Francisco Adan Gil (CS Alfaró, La Rioja); Emilio García Criado (Centro de Salud del Carpio, Córdoba); Rafael Durá Belinchón (CS Godella, Valencia); Jose M. Fernández Toro (CS Zona Centro, Cáceres); Juan A. Divisón Garrote (Centro de Salud de Casas Ibáñez, Consultorio de Fuentealbilla, Albacete).

Supplementary tables

Table S1. Minimum detectable, intra- and inter-assay variability of kits used

Biomarkers	Minimum detectable	Intra-assay variability, %	Inter-assay variability, %
Eotaxin, pg/mL	4	7.2	10.8
FGF-2, pg/mL	7.6	2.3	4.8
Fractalkine, pg/mL	22.7	4.5	9.4
GM-CSF, pg/mL	7.5	3.1	10.1
IFN- γ , pg/mL	0.8	1.6	12
IP-10, pg/mL	8.3	2.6	15.3
Leptin, pg/mL	7.3	<10	<15
MCP-1, pg/mL	1.9	1.5	7.9
MDC, pg/mL	3.6	1.6	7.2
MIP-1 β , pg/mL	3	2.4	8.8
MMP-9, pg/mL	2	1.9	9
MPO, ng/mL	0.036	<15	<20
OC, pg/mL	68.5	<10	<15
OPG, pg/mL	1.9	<10	<15
OPN, pg/mL	37.7	<10	<15
PAI-1, pg/mL	3.2	1.2	9.9
sICAM-1, pg/mL	0.019	<15	<20
sVCAM-1, pg/mL	0.024	<15	<20
VEGF, pg/mL	26.3	3.7	10.4

FGF-2: fibroblast growth factor 2; GM-CSF: granulocyte-macrophage colony-stimulating factor; IFN- γ : interferon gamma; IP-10: IFN- γ inducible protein 10; MCP-1: monocyte chemoattractant protein-1; MDC: macrophage-derived chemokine; MIP-1 β : macrophage inflammatory protein-1beta; MMP-9: matrix metalloproteinase-9; MPO: myeloperoxidase; OC: osteocalcin; OPG: osteoprotegerin; OPN: osteopontin; PAI-1: plasminogen activator inhibitor-1; sICAM-1: soluble intercellular adhesion molecule-1; sVCAM-1: soluble vascular cell adhesion molecule-1; VEGF: vascular endothelial growth factor.

Table S2. Variable importance of the random survival forest analysis with all variables

Biomarkers	100 × VIMP (95% CI)	5th percentile*
OPN	1.214 (0.823, 1.606)	0.886
PP	1.087 (0.736, 1.437)	0.793
SBP	1.048 (0.710, 1.385)	0.764
Diabetes	0.769 (0.521, 1.018)	0.561
Lipid-lowering drugs	0.614 (0.416, 0.811)	0.448
OPG	0.576 (0.390, 0.762)	0.420
DBP	0.524 (0.355, 0.693)	0.382
Atrial fibrillation	0.491 (0.332, 0.649)	0.358
CKD stage	0.488 (0.331, 0.646)	0.356
Antihypertensive drugs	0.455 (0.308, 0.602)	0.332
Anti-diabetic drugs	0.441 (0.299, 0.583)	0.322
HDL-cholesterol	0.412 (0.279, 0.545)	0.301
Hypertension	0.411 (0.279, 0.544)	0.300
OC	0.351 (0.238, 0.465)	0.256
25(OH)-vitamin D3	0.343 (0.232, 0.453)	0.250
Age	0.298 (0.202, 0.394)	0.217
Dyslipidemia	0.266 (0.180, 0.352)	0.194
PAI-1	0.261 (0.177, 0.345)	0.190
Presence of plaque at baseline	0.246 (0.166, 0.325)	0.179
BMI	0.241 (0.163, 0.318)	0.176
LDL-cholesterol	0.233 (0.158, 0.308)	0.170
Total cholesterol	0.224 (0.152, 0.296)	0.163
Potassium	0.214 (0.145, 0.283)	0.156
sVCAM-1	0.175 (0.119, 0.231)	0.128
Fractalkine	0.170 (0.115, 0.225)	0.124
hsCRP	0.157 (0.106, 0.208)	0.115
MPO	0.143 (0.097, 0.190)	0.105
Eotaxin	0.127 (0.086, 0.168)	0.093
Leptin	0.126 (0.085, 0.167)	0.092
Calcium	0.112 (0.076, 0.149)	0.082
MMP-9	0.107 (0.073, 0.142)	0.078
Heart failure	0.101 (0.069, 0.134)	0.074
Triglycerides	0.100 (0.068, 0.132)	0.073
IFN- γ	0.053 (0.036, 0.070)	0.039
VEGF	0.053 (0.036, 0.070)	0.039
FGF-2	0.043 (0.029, 0.057)	0.031
Phosphorus	0.016 (0.011, 0.021)	0.012
MCP-1	0.011 (0.007, 0.014)	0.008
sICAM-1	-0.004 (-0.005, -0.003)	-0.005
Smoking habits	-0.005 (-0.007, -0.004)	-0.007
MDC	-0.052 (-0.068, -0.035)	-0.066
IP-10	-0.061 (-0.081, -0.042)	-0.078

Continued on next page...

...Continued from previous page

Biomarkers	100 × VIMP (95% CI)	5th percentile*
Women	-0.067 (-0.089, -0.046)	-0.085
GM-CSF	-0.072 (-0.095, -0.049)	-0.092
Family history of early CVD	-0.084 (-0.112, -0.057)	-0.107
MIP-1 β	-0.086 (-0.114, -0.058)	-0.109

* With purpose of selecting variables: 5th percentile > 0 (in bold).

BMI: body mass index; CVD: cardiovascular disease; DBP: diastolic blood pressure; FGF-2: fibroblast growth factor 2; GM-CSF: granulocyte-macrophage colony-stimulating factor; HDL: high-density lipoprotein; hsCRP: high-sensitivity C-reactive protein; IFN- γ : interferon gamma; IP-10: IFN- γ inducible protein 10; LDL: low-density lipoprotein; MCP-1: monocyte chemoattractant protein-1; MDC: macrophage-derived chemokine; MIP-1 β : macrophage inflammatory protein-1beta; MMP-9: matrix metalloproteinase-9; MPO: myeloperoxidase; OC: osteocalcin; OPG: osteoprotegerin; OPN: osteopontin; PAI-1: plasminogen activator inhibitor-1; PP: pulse pressure; SBP: systolic blood pressure; sICAM-1: soluble intercellular adhesion molecule-1; sVCAM-1: soluble vascular cell adhesion molecule-1; VEGF: vascular endothelial growth factor.

3.4 Mètodes i resultats de l'objectiu 4

En aquesta secció es presenten els mètodes utilitzats i els resultats obtinguts per donar resposta al quart dels objectius plantejats en aquesta tesi. Tota la metodologia i els resultats estan continguts en l'esborrany del manuscrit original Forné *et al.* (actualment en procés de revisió), la informació del qual es detalla a continuació, junt amb el resum traduït al català.

Títol: A model-based cost-utility analysis of increasing percutaneous coronary intervention use in elderly patients with acute coronary syndromes

Autors: Forné C, Subirana I, Blanch J, Ferrières J, Azevedo A, Meisinger C, Farmakis D, Tavazzi L, Davoli M, Ramos R, Brosa M, Marrugat J, Décano IR; on behalf of the EUROTRACS investigators

Revista: *manuscrit en versió esborrany*

Factor d'impacte de l'any de la publicació: *manuscrit en versió esborrany*

Resum:

Antecedents: La intervenció coronària percutània (ICP) redueix la mortalitat en pacients amb síndrome coronària aguda (SCA), però es desconeix l'eficiència d'augmentar-ne l'ús en pacients ancians amb SCA.

Mètodes: Es va avaluar l'eficiència de l'augment de l'ús de la ICP en comparació amb la pràctica actual en pacients amb ≥ 75 anys ingressats per SCA a França, Alemanya, Grècia, Itàlia, Portugal i Espanya amb un model de semi-Markov de transició entre estats. Les estimacions de la reducció de la mortalitat hospitalària per ús de la ICP i dels costos van ser obtingudes del projecte EUROTRACS EU (n=28.600). El risc de recurrència i de mort extrahospitalària per totes les causes es van obtenir de la base de dades SIDIAP del nord-est d'Espanya (n=55.564). Es va modelitzar la mortalitat hospitalària mitjançant l'anàlisi estratificada del *propensity score* (PS). Es va estimar el risc de recurrència d'SCA i la mortalitat extrahospitalària en un horitzó temporal de 8 anys ajustant un model de supervivència multiestat. Els escenaris analitzats van ser augmentar l'ús de la ICP entre pacients amb les probabilitats més altes, moderades i més baixes de rebre ICP basades en l'anàlisi de PS.

Resultats: França, Grècia i Portugal van mostrar uns costos totals similars (7,29-11,05 milions d'euros per cada 1.000 individus), mentre que a Alemanya, Itàlia i Espanya, els costos van ser més alts (13,53-22,57 milions d'euros per cada 1.000 individus). Les relacions de cost-utilitat incremental de proporcionar ICP a tots els pacients oscil·laven entre 2.262,8€ per any de vida ajustat per qualitat (AVAQ) guanyat, per als homes alemanys, i 6.324,3€/AVAQ guanyat, per a les dones italianes. L'augment de l'ús de la ICP es va mostrar eficient en un llindar de disposició a pagar de 10.000€/AVAQ guanyat, per a tots els escenaris en els sis països, en homes i dones.

Conclusions: En comparació amb la pràctica clínica actual, l'ampliació de l'ús de la ICP en pacients ancians amb SCA seria eficient en diferents sistemes de salut a Europa, independentment de l'estratègia seleccionada.

A model-based cost-utility analysis of increasing percutaneous coronary intervention use in elderly patients with acute coronary syndromes

Carles Forné ^a, Isaac Subirana ^{b,c}, Jordi Blanch ^d, Jean Ferrières ^e, Ana Azevedo ^f, Christa Meisinger ^{g,h}, Dimitrios Farmakis ⁱ, Luigi Tavazzi ^j, Marina Davoli ^k, Rafel Ramos ^{d,l-n}, Max Brosa ^o, Jaume Marrugat ^{b,p}, Irene Román Dégano ^{b,p,q,*}; EUROTRACS Investigators

^a Department of Basic Medical Sciences, University of Lleida, Lleida, Spain; ^b REGICOR Study Group, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain; ^c CIBER Epidemiology and Public Health, ISCIII, Madrid, Spain; ^d Vascular Health Research Group (ISV-Girona), Jordi Gol Institute for Primary Care Research (IDIAPJGol), Barcelona, Spain; ^e Department of Cardiology, Toulouse Rangueil Hospital, Toulouse University School of Medicine, Toulouse, France; ^f Department of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School and EPIUnit Institute of Public Health of the University of Porto (ISPUP), Porto, Portugal; ^g MONICA/KORA Myocardial Infarction Registry, University Hospital of Augsburg, Augsburg, Germany; ^h Helmholtz Zentrum München, German Research Center for Environmental Health, Independent Research Group Clinical Epidemiology, Neuherberg, Germany; ⁱ Second Department of Cardiology, Attikon University Hospital, National and Kapodistrian, University of Athens Medical School, Athens, Greece; ^j Maria Cecilia Hospital, GVM Care & Research, Cotignola, Italy; ^k Department of Epidemiology, Lazio Regional Health Service, Rome, Italy; ^l Catalan Institute of Health, Barcelona, Spain; ^m Department of Medical Sciences, School of Medicine, University of Girona, Girona, Spain; ⁿ Girona Biomedical Research Institute (IdIBGi), Girona, Spain; ^o Oblikue Consulting SL, Barcelona, Spain; ^p CIBERCV of Cardiovascular Diseases, Instituto de Salud Carlos III (ISCIII), Madrid, Spain; ^q Faculty of Medicine, University of Vic-Central University of Catalonia (UVic-UCC), Vic, Spain.

A full roster of EUROTRACS investigators can be found at www.regicor.org/eurotracs_inv

* Address for correspondence: Irene Román Dégano (iroman@imim.es), REGICOR Study Group, IMIM, Barcelona Biomedical Research Park (PRBB). Dr. Aiguader 88, 08003, Barcelona, Spain. Phone: +34 93 3160714, Fax: +34 93 3160796

Abstract

Background: Percutaneous coronary intervention (PCI) reduces mortality in acute coronary syndrome (ACS) patients but the cost-utility of increasing its use in elderly ACS patients is unknown.

Methods: We assessed the efficiency of increased PCI use compared to current practice in patients aged ≥ 75 years admitted for ACS in France, Germany, Greece, Italy, Portugal, and Spain with a semi-Markov state transition model. In-hospital mortality reduction estimates by PCI use and costs were derived by the EUROTRACS EU Project (n=28,600). Risk of recurrence and out-of-hospital all-cause mortality were obtained from the SIDIAP database from North-Eastern Spain (n=55,564). In-hospital mortality was modelled using stratified propensity score (PS) analysis. 8-year ACS recurrence risk and out-of-hospital mortality were estimated with a multistate survival model. The scenarios analyzed were to increase PCI use among patients with the highest, moderate, and lowest probability of receiving PCI based on the PS analysis.

Results: France, Greece, and Portugal showed similar total costs/1000 individuals (7.29-11.05 million €); while in Germany, Italy, and Spain, costs were higher (13.53-22.57 million €). Incremental cost-utility ratios of providing PCI to all patients ranged from 2,262.8€ per quality adjusted life year (QALY) gained for German males to 6,324.3€/QALY gained for Italian females. Increasing PCI use was cost-effective at a willingness-to-pay threshold of 10,000€/QALY gained for all scenarios in the 6 countries, in males and females.

Conclusions: Compared to current clinical practice, broadening PCI use in elderly ACS patients would be cost-effective across different healthcare systems in Europe, regardless of the selected strategy.

Keywords: Elderly; Acute coronary syndrome; Percutaneous coronary intervention; Cost-effectiveness; Cost-utility.

Introduction

Seventeen to 34% of the patients admitted for acute coronary syndromes (ACS) are elderly^{1,2} and their management should be similar to that of younger patients. In patients with ST-segment elevation myocardial infarction (STEMI), the standard of care is immediate reperfusion.³ Primary percutaneous coronary intervention (PCI) is the preferred reperfusion strategy with no upper age limit. In patients with non-ST-elevation ACS (NSTEMI-ACS), an invasive strategy is recommended if patients have an intermediate/high risk,⁴ which represents most NSTEMI-ACS patients. For high-risk patients, the standard of care is a routine invasive strategy. In elderly patients, the routine compared to selective invasive strategy reduces the occurrence of acute myocardial infarction (AMI) and of subsequent revascularization.⁵

Although current guidelines recommend an invasive strategy in all STEMI and most NSTEMI-ACS patients, population-based registries have described lower PCI and angiography use in elderly ACS patient.^{6,7} The lower use of evidence-based therapies in elderly ACS patients has been associated with presence of comorbidities and occurrence of atypical symptoms.⁸ In addition, elderly ACS patients have an increased risk of thrombotic and bleeding events, and also of PCI complications.^{1,9–12} On the other hand, it is possible that part of the underutilization of evidence-based therapies is caused by the unknown cost-utility of increasing their performance in elderly ACS patients.

To fill this gap, we carried out the EUROTRACS project,¹³ which was focused on examining the cost-utility of different interventions to improve cardiovascular disease outcomes in Europe. The aim of this study was to examine the efficiency of a broader use of PCI in the management of elderly patients admitted for ACS in six European countries.

Methods

Model description

A semi-Markov state transition model was developed to assess the efficiency of increased PCI use in ACS patients aged ≥ 75 years. The model conceptualized the course of the disease in terms of mutually exclusive health states and transitions among them. The health states considered were ACS hospitalized event—the starting health state for all patients—eventual recurrences during the subsequent years after an AMI or unstable angina (UA), and death (Supplementary Figure 1). The model used yearly cycles, except for the first year that was divided in two periods—in-hospital stay and the rest of year.

Model inputs

Model inputs included estimates of in-hospital mortality, PCI use, transition probabilities, complications during hospitalization, costs, and utility decrements. In-hospital mortality reductions associated to PCI utilization were obtained from a tertile stratified propensity score (PS) analysis carried out

with ACS patient data from the EUROTRACS project.¹³ Patients were included if they had an AMI or UA, were aged ≥ 75 years and had information in the variables required for the analysis ($n=28,600$) (Supplementary Figure 2). Missing data in the required variables ranged from 0.05 to 4.63%. Characteristics of the ACS patients from the EUROTRACS ACS cohort are presented in Supplementary Table 1-A. The PS was defined as the probability of receiving PCI and estimated using a logistic regression model where the predictor variables were patient characteristics at hospital admission associated with PCI and with in-hospital mortality (age, diabetes, hypertension, previous MI, admission Killip class, and ST-segment elevation). The area under the ROC curve of the PS logistic model was approximately 70% in males and in females. The association between in-hospital mortality and PCI was assessed separately for each PS tertile and sex using logistic regression adjusted for patient characteristics during hospitalization associated with mortality (reception of PCI, PS, age, reception of thrombolysis, maximum Killip during hospitalization, and ACS type [AMI/UA]). Markov model inputs included in-hospital mortality reductions by sex and by PS tertile, obtained by subtracting the mortality rate in patients receiving PCI from the rate in those not receiving PCI. Rates of PCI use and in-hospital mortality rates in elderly EUROTRACS patients are shown in Supplementary Table 2.

Transition probabilities between health states were obtained by fitting a multistate survival model in a population-based cohort with an 8-year follow-up. Patients with primary ACS during 2007-2013, were selected from the Information System for the Development of Research in Primary Care (SIDIAP), a validated healthcare database from North-Eastern Spain.^{14,15} Patients were included if they had a primary AMI or UA, were aged ≥ 30 years, and did not die during the first month after diagnosis ($n=55,564$) (Supplementary Figure 3). There were no missing data and individuals lost to follow-up were censored (1,050/55,564). Characteristics of the ACS patients from the SIDIAP cohort are presented in Supplementary Table 1-B. Each participant was followed-up until death, or censoring applied to transfer from SIDIAP or end of study period, whichever occurred first. Deaths during the first month after the ACS were excluded to avoid overestimation of in-hospital mortality, already accounted with the stratified PS analysis. We used the illness-death without recovery model, a three-state model with three transitions. Patients entered the 1st state when they were discharged alive from their first hospitalization for AMI or UA; the 2nd state was a recurrence, a secondary ACS event; and the 3rd state was all-cause mortality. The predictor variables for transition hazards were age, sex, reception of PCI, and ACS type (AMI/UA) (Supplementary Table 3). Possible trajectories—paths that a patient might follow—through the multistate model were sampled using 10,000 bootstrap data sets for computing prediction probabilities (Supplementary Figure 4). The multistate survival model was developed with the R packages *mstate* and *survival*.

Cost data

Each event and health state of the Markov model was associated with defined country-specific costs obtained from the EUROTRACS project (Supplementary Table 4). Acute care cost of surviving AMI/UA patients were considered as weighted cost of patients with or without complications during hospitalization. Percentage of complications during hospitalization was assumed to be 45.9% according to hospital activity from the Spanish National Health System.¹⁶ Both acute care and follow-up costs varied extensively between participating countries.

Health-related quality of life

The impact of increased PCI use on mortality and quality-of-life was estimated using quality-adjusted life years (QALYs). QALYs were derived by applying disutility values to the estimated years of life. Health-related utility decrements were assigned to each health state. ACS-specific disutilities were adapted from previous studies,^{17,18} For death, we assumed a disutility of 1. Disutility values for AMI/UA events in the first or subsequent years were -0.1055 and -0.1 , respectively.

Perspective, time horizon and discounting

The analysis was carried out from the perspective of the healthcare payer, so it considered only health effects for the patients and direct costs faced by the healthcare provider. The time horizon was 8 years, and all costs were expressed as year 2018 EUR. Costs and QALYs were discounted using an annual rate of 3%.

Scenarios compared

Four scenarios were compared to the current PCI utilization—base scenario. A: PCI use in patients included in the PS third tertile, which had the highest probability of receiving PCI; B: PCI use in patients included in the PS second and third tertiles, which had a moderate-high probability of receiving PCI; C: PCI use in all patients; and D: PCI use in patients included in the PS first tertile, which had the lowest probability of receiving PCI.

We analyzed the 4 scenarios in a cohort of initial age 75 years, and considering females and males separately.

Analysis

Cost-utility of increasing PCI use was expressed as the incremental cost-utility ratio (ICUR), which describes the incremental cost per unit of QALY gained. ICURs were calculated for each scenario compared to the base one, as follows:

$$\frac{\text{Cost}_{\text{alternative scenario}} - \text{Cost}_{\text{base scenario}}}{\text{QALY}_{\text{alternative scenario}} - \text{QALY}_{\text{base scenario}}}$$

We performed a deterministic one-way sensitivity analysis to examine the impact on results of input changes. Input values for each parameter were varied one at a time across a range while the remaining values were held at their baseline values (Supplementary Table 5).

We did a probabilistic sensitivity analysis (PSA) to explore the combined impact of all uncertainties in model inputs. We specified each input as a probability distribution (Supplementary Table 6). Monte Carlo simulation was used to generate 1,000 samples in the PSA. PSA results were presented with incremental scatterplots (cost-utility planes) and with cost-utility acceptability curves for each alternative scenario and country.

Analysis of in-hospital mortality, transition probabilities, and PSA graphs, were performed with the R statistical software (version 3.5.2). The semi-Markov model and the sensitivity analyses were developed in MS Excel Spreadsheet 2013 with a specific macro.

Results

Receiving PCI was significantly associated with a lower risk of in-hospital mortality in all PS groups (Supplementary Table 7).

Table 1 shows the cost-utility results for each participant country by sex. Providing PCI to all patients in the third PS tertile (scenario A) was associated with the lowest increase in QALYs at the lowest cost, while providing PCI to all patients (scenario C) was associated with the highest increase in QALYs at the highest cost. Total costs and QALYs were similar between sexes. Total QALYs/1000 individuals ranged from 4,795.2 to 5,439.9. Total costs between scenarios were similar within countries but different between countries. France, Greece, and Portugal showed similar total costs/1000 individuals (7.29-11.05 million €) depending on the scenario; while in Germany, Italy, and Spain, costs were 2-3 times higher (13.53-22.57 million €). ICUR ranged from 2,262.8 €/QALY gained for German males in scenario C to 6,422.4 €/QALY gained for Italian females in scenario B. Regardless of the country, all scenarios provided similar efficiency results, especially for females. As for results by sex, ICURs were higher in females than in males for all scenarios and countries, particularly in the scenarios of providing PCI to all patients (scenario C) and to the ones in the first PS tertile (scenario D).

Results were robust across a wide range of values for the model parameters as shown in the deterministic one-way sensitivity analysis (Supplementary Tables 8-13). The parameters with the largest influence were disutilities, cost of PCI, in-hospital mortality reduction by PCI, and annual discount rate, for all countries and regardless of sex. Results of PSA showed that all alternative scenarios yielded more QALYs, at a higher cost, compared to current clinical practice (Figure 1). ICURs for increasing PCI use among elderly ACS patients fell below a willingness-to-pay of 10,000 €/QALY gained in almost 100% of the PSA samples, in both males and females (Figure 2). All scenarios showed

Table 1. Results of economic evaluation: total cost and quality adjusted life year (QALY) for each scenario, and incremental cost-utility ratios (ICUR) by country and sex

Country	Sex		Scenarios				
			Base	A	B	C	D
France	Females	Total cost	8.00	8.70	9.60	10.66	9.06
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	4,335.5	4,583.4	4,479.7	4,331.2
	Males	Total cost	8.29	8.88	9.65	10.66	9.30
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	3,951.3	4,190.4	3,887.2	3,543.3
Germany	Females	Total cost	14.84	15.29	15.86	16.53	15.52
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	2,756.6	2,890.4	2,833.0	2,750.9
	Males	Total cost	15.46	15.81	16.25	16.84	16.06
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	2,293.8	2,423.2	2,262.8	2,081.0
Greece	Females	Total cost	8.81	9.40	10.15	11.05	9.70
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	3,656.9	3,829.3	3,757.3	3,654.3
	Males	Total cost	8.81	9.31	9.94	10.81	9.68
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	3,328.8	3,490.2	3,285.2	3,052.7
Italy	Females	Total cost	18.80	19.80	21.05	22.57	20.32
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	6,188.2	6,422.4	6,324.3	6,183.9
	Males	Total cost	18.84	19.68	20.73	22.23	20.34
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	5,606.8	5,821.2	5,549.9	5,242.2
Portugal	Females	Total cost	7.29	7.74	8.31	8.99	7.97
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	2,774.8	2,917.5	2,857.4	2,771.4
	Males	Total cost	7.64	8.01	8.49	9.14	8.28
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	2,494.2	2,632.3	2,458.2	2,260.8
Spain	Females	Total cost	13.53	14.40	15.51	16.84	14.85
		Total QALY	4,795.2	4,957.3	5,145.7	5,390.7	5,040.2
		ICUR	-	5,420.3	5,671.5	5,566.0	5,415.0
	Males	Total cost	13.98	14.72	15.66	16.95	15.27
		Total QALY	4,829.6	4,980.2	5,153.9	5,439.9	5,115.6
		ICUR	-	4,928.3	5,168.2	4,865.1	4,521.4

Results from a hypothetical cohort of 1,000 individuals with primary ACS at 75 years old and an 8-year follow-up. Absolute values have been discounted at an annual rate of 3%. Total cost is in millions of euros. Scenario A: use of PCI in all patients included in the third tertile of the PS analysis (highest probability of receiving PCI); Scenario B: use of PCI in all patients included in the second and third tertiles of the PS analysis (moderate-high probability of receiving PCI); Scenario C: use of PCI in all patients with ACS—regardless of their PS tertile—; Scenario D: use of PCI in patients included in the first tertile of the PS analysis (lowest probability of receiving PCI). ACS, acute coronary syndrome; ICUR, incremental cost-utility ratio; PCI, percutaneous coronary intervention; PS, propensity score; QALY, quality-adjusted life year.

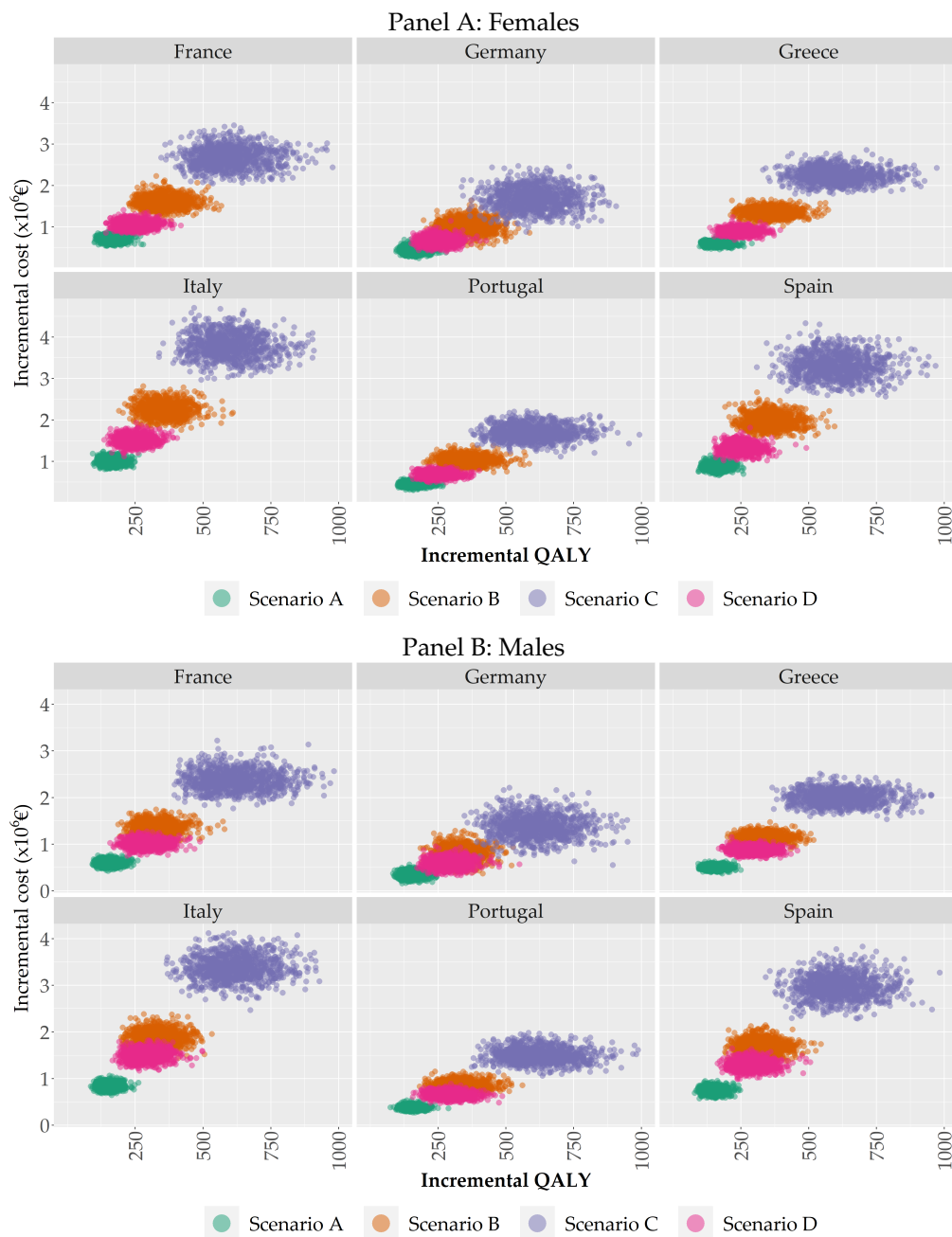


Figure 1. Cost-utility planes by sex and country. Scatterplot showing the mean differences in costs and in QALYs. For each alternative scenario, estimates of incremental costs and QALYs were obtained by re-running the model employing 1,000 bootstrap samples from the related probability distributions in the probabilistic sensitivity analysis. Scenario A: use of PCI in all patients included in the third tertile of the PS analysis (highest probability of receiving PCI); Scenario B: use of PCI in all patients included in the second and third tertiles of the PS analysis (moderate-high probability of receiving PCI); Scenario C: use of PCI in all patients with ACS—regardless of their PS tertile—; Scenario D: use of PCI in patients included in the first tertile of the PS analysis (lowest probability of receiving PCI). ACS, acute coronary syndrome; PCI, percutaneous coronary intervention; PS, propensity score; QALY, quality-adjusted life year.

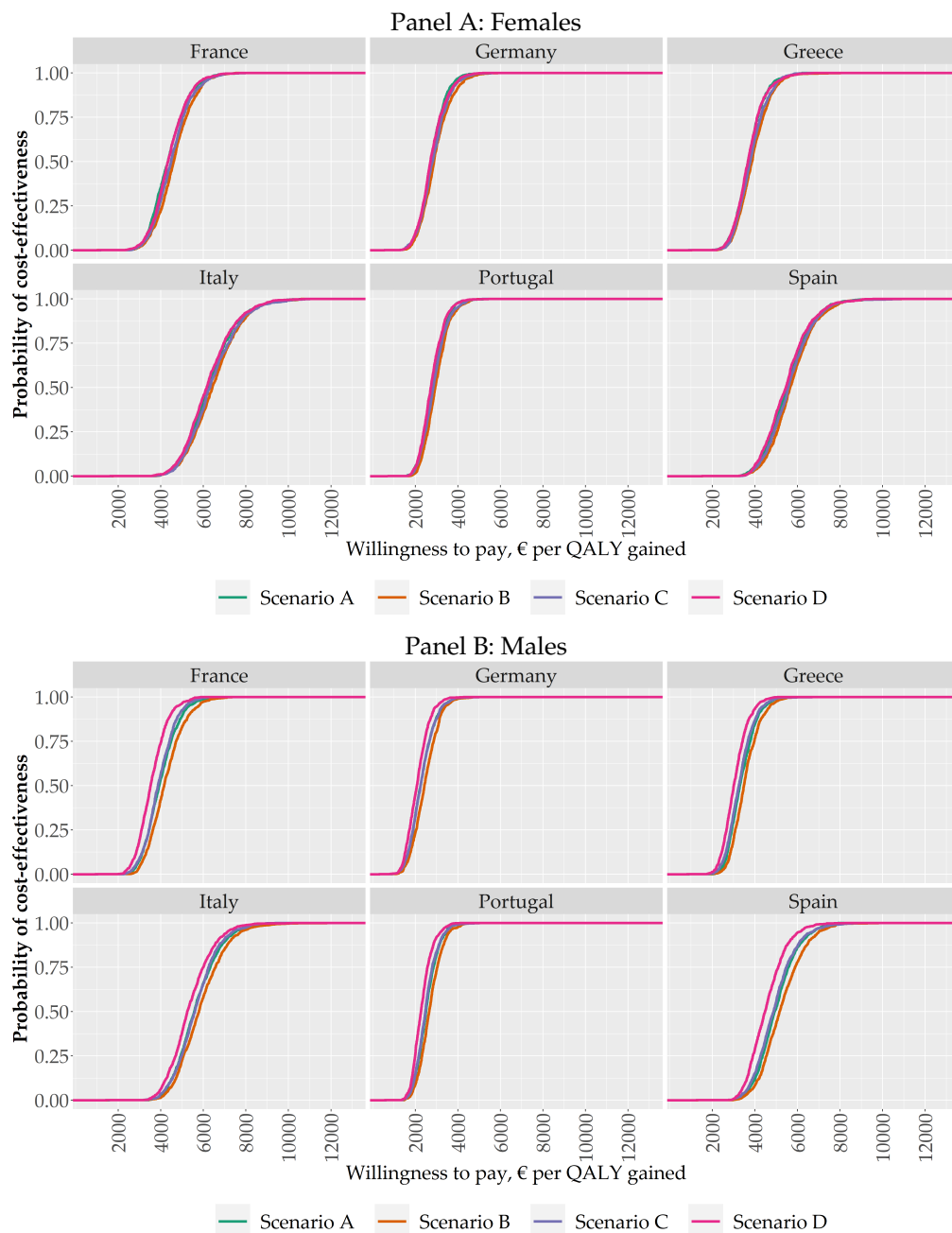


Figure 2. Cost-utility acceptability curves by sex and country. Cost-utility acceptability curves presenting the probability that each alternative scenario is optimal given different values of willingness to pay for an additional QALY, derived from 1,000 bootstrap samples in the probabilistic sensitivity analysis. Scenario A: use of PCI in all patients included in the third tertile of the PS analysis (highest probability of receiving PCI); Scenario B: use of PCI in all patients included in the second and third tertiles of the PS analysis (moderate-high probability of receiving PCI); Scenario C: use of PCI in all patients with ACS—regardless of their PS tertile—; Scenario D: use of PCI in patients included in the first tertile of the PS analysis (lowest probability of receiving PCI). ACS, acute coronary syndrome; PCI, percutaneous coronary intervention; PS, propensity score; QALY, quality-adjusted life year.

a higher probability of being cost-effective in males than in females, for the same willingness-to-pay threshold.

Discussion

In this study with an 8-year time horizon, increasing the number of elderly ACS patients who received a PCI was cost-effective at a willingness-to-pay threshold of 10,000 €/QALY gained in all 6 European countries studied.

While all scenarios provided similar efficiency results, offering PCI to all ACS patients with the lowest probability of receiving a PCI, was the most cost-effective in males. In females, cost-utility results were almost the same for providing PCI to all ACS patients with the lowest and the highest probability of receiving a PCI. This result reflects the observed in-hospital mortality reductions by PCI use which were larger in males in the first PS tertile, and in females in the first and third PS tertile. It also suggests that the impact of PCI use on health outcomes is larger in the in-hospital period than in the following 8-years.¹⁹

The fact that one of the most cost-effective scenarios was to provide PCI to all ACS patients with the lowest probability of receiving PCI may indicate that the decision of providing PCI depends on many factors, some of which were probably not included in this study and cannot always be assessed. We could expect that a fraction of patients classified with a low probability of receiving PCI could have indeed a higher probability of receiving PCI. The PS model used to calculate the probability of receiving PCI was based on age, type of ACS, occurrence of previous AMI, and presence of diabetes, hypertension, and heart failure. The PS model had a discrimination of 70% showing that the indication of PCI could not be totally defined with the available covariates. Current clinical guidelines on STEMI and NSTEMI-ACS recommend evaluating comorbidities, life expectancy, quality of life, frailty and patient preferences, to indicate an invasive strategy in elderly patients^{3,4} Comorbidities are common in elderly ACS patients and are associated with worse prognosis,²⁰ but as the benefit of an invasive strategy is not fully understood, these patients have a lower probability of receiving an invasive strategy.²¹ Similarly, frailty is associated with higher in-hospital and long-term mortality in NSTEMI-ACS patients but frail patients have a lower probability of receiving PCI.²² These studies and our results suggest that some high-risk ACS patients are probably treated more conservatively than they should.

Total costs associated with short- and long-term management of ACS patients, and therefore ICUR by increasing the number of patients receiving PCI, were different among countries. Previous studies have also shown large differences in the cost of cardiovascular events across countries, mainly due to differences in study population and costing methodologies.^{23–27} While cost differences were associated with a reduced cost-utility in those countries with higher unit costs, it is important to bear in mind that ICUR was lower than 10,000 €/QALY gained for males and females in all countries.

Our model showed a high structural robustness, reflected in the consistency of the results for the different scenarios considered. It also provided very reliable results, probably due to the use of large cohorts of ACS patients to obtain the model parameters, and in spite of the wide ranges considered for parameters with unknown dispersion. Nevertheless, predicted probabilities of the possible paths through the multistate model were not included in the sensitivity analysis, as they were obtained by intensive computational bootstrapping methods, that made its inclusion in the sensitivity analysis unfeasible.

There were other limitations in this study that should be considered. First, we were unable to analyze the cost-utility for ACS patients with and without ST-segment elevation because this information was not available in the SIDIAP database.^{14,15} However, taking into account that all STEMI and most NSTEMI-ACS patients are candidates to receive a PCI, this analysis provides a robust first picture of the cost-utility of increasing PCI use in elderly ACS patients. Second, the study population may not reflect current ACS patients and practice. While PCI rate and in-hospital mortality estimates were obtained from a large database of ACS patients in Europe (EUROTRACS),¹³ not all patients were admitted during the last decade. Although we would expect an increased PCI rate and a reduction in in-hospital mortality in more contemporary data, PCI rate has not been increased in elderly ACS patients at the same rate as in younger patients.²⁸ Third, short- and long-term recurrence data were obtained from the validated SIDIAP database and these data was assumed to be similar for the other 5 countries. It is known that there are differences in clinical management among European countries.²⁹ However, it could be expected that the natural history of the disease would be similar among countries.

Conclusions

Compared to current clinical practice, broadening PCI use in elderly ACS patients would be cost-effective across different healthcare systems in Europe, regardless of the selected strategy. Cost-utility should not be a factor for not providing PCI to elderly ACS patients. The fact that providing PCI to elderly ACS patients with low probability of receiving a PCI is as cost-effective as providing PCI to patients with high probability of receiving a PCI, suggests that the criteria to recommend PCI in elderly ACS patients should be more clearly defined.

Author contributions

CF, IRD, JM and MB contributed to the conception or design of the work. All authors contributed to the acquisition, analysis, or interpretation of data for the work. CF and IRD drafted the manuscript. All authors critically revised the manuscript. All gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of conflicting of interests

Authors declare no conflict of interest.

Acknowledgments

This article will be part of the doctoral thesis of Carles Forné and is within the doctoral program in Health of the University of Lleida.

Funding

This work was supported by the Consumers, Health, Agriculture and Food executive agency of the European Commission (EC) [EUROTRACS 2012 / 1207]; the Spanish Ministry of Science and Innovation, co-financed with European Union European Regional Development Funds—ERDF—[CIBER of Cardiovascular Diseases CB16/11/00229 and 00246, initiatives of the Instituto de Salud Carlos III]; the Health Department of the Generalitat de Catalunya through the Agència de Gestió d'Ajuts Universitaris de Recerca de Catalunya (AGAUR) [2017SGR222], and the Health Research and Innovation Strategic Plan (PERIS) [SLT006/17/00029 to IRD].

References

1. Rosengren A, Wallentin L, Simoons M et al. Age, clinical presentation, and outcome of acute coronary syndromes in the Euroheart acute coronary syndrome survey. *Eur Heart J*. 2006;27:789-795.
2. Benjamin EJ, Blaha MJ, Chiuve SE et al. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*. 2017;135:e146-e603.
3. Ibanez B, James S, Agewall S et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2018;39:119-177.
4. Neumann FJ, Sousa-Uva M, Ahlsson A et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J*. 2019;40:87-165.
5. Garg A, Garg L, Agarwal M et al. Routine Invasive Versus Selective Invasive Strategy in Elderly Patients Older Than 75 Years With Non-ST-Segment Elevation Acute Coronary Syndrome: A Systematic Review and Meta-Analysis. *Mayo Clin Proc*. 2018;93:436-444.
6. Schoenenberger AW, Radovanovic D, Windecker S et al. Temporal trends in the treatment and outcomes of elderly patients with acute coronary syndrome. *Eur Heart J*. 2016;37:1304-1311.
7. Malta Hansen C, Wang TY, Chen AY et al. Contemporary Patterns of Early Coronary Angiography Use in Patients With Non-ST-Segment Elevation Myocardial Infarction in the United States: Insights From the

- National Cardiovascular Data Registry Acute Coronary Treatment and Intervention Outcomes Network Registry. *JACC Cardiovasc Interv.* 2018;11:369-380.
8. Madhavan MV, Gersh BJ, Alexander KP, Granger CB and Stone GW. Coronary Artery Disease in Patients ≥ 80 Years of Age. *J Am Coll Cardiol.* 2018;71:2015-2040.
 9. Fox KA, Clayton TC, Damman P et al. Long-term outcome of a routine versus selective invasive strategy in patients with non-ST-segment elevation acute coronary syndrome a meta-analysis of individual patient data. *J Am Coll Cardiol.* 2010;55:2435-2445.
 10. Valgimigli M, Bueno H, Byrne RA et al. 2017 ESC focused update on dual antiplatelet therapy in coronary artery disease developed in collaboration with EACTS: The Task Force for dual antiplatelet therapy in coronary artery disease of the European Society of Cardiology (ESC) and of the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J.* 2018;39:213-260.
 11. Skolnick AH, Alexander KP, Chen AY et al. Characteristics, management, and outcomes of 5,557 patients age $>$ or $=90$ years with acute coronary syndromes: results from the CRUSADE Initiative. *J Am Coll Cardiol.* 2007;49:1790-1797.
 12. Ndrepepa G, Neumann FJ, Schulz S et al. Incidence and prognostic value of bleeding after percutaneous coronary intervention in patients older than 75 years of age. *Catheter Cardiovasc Interv.* 2014;83:182-189.
 13. Dégano IR, Subirana I, Fusco D et al. Percutaneous coronary intervention reduces mortality in myocardial infarction patients with comorbidities: Implications for elderly patients with diabetes or kidney disease. *Int J Cardiol.* 2017;249:83-89.
 14. Bolívar B, Fina Avilés F, Morros R et al. [SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research]. *Med Clin (Barc).* 2012;138:617-621.
 15. Ramos R, Balló E, Marrugat J et al. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol (Engl Ed).* 2012;65:29-37.
 16. Ministry of Health, Consumer Affairs and Social Welfare [Ministerio de Sanidad, Consumo y Bienestar Social]. Minimum Basic Data Set-Hospital Discharge [Conjunto Mínimo Básico de Datos - Hospitalización, CMBD-H]. Available at: <https://pestadistico.inteligenciadegestion.mscbs.es>. Accessed Nov 10, 2019.
 17. Kourlaba G, Fragoulakis V, Maniadakis N. Economic evaluation of clopidogrel in acute coronary syndrome patients without ST-segment

- elevation in Greece: a cost-utility analysis. *Appl Health Econ Health Policy*. 2012;10:261-271.
18. Lindgren P, Kahan T, Poulter N et al. Utility loss and indirect costs following cardiovascular events in hypertensive patients: the ASCOT health economic substudy. *Eur J Health Econ*. 2007;8:25-30.
 19. Khan M, Kayani W, Dadu R et al. Short- and long-term outcomes of percutaneous coronary intervention versus coronary artery bypass grafting in older (>70 years) patients with unprotected left main coronary artery disease in the era of drug eluting stents. *JACC* 2016;67(13):Supplement.
 20. Sanchis J, Núñez J, Bodí V et al. Influence of comorbid conditions on one-year outcomes in non-ST-segment elevation acute coronary syndrome. *Mayo Clin Proc*. 2011;86:291-296.
 21. Fox KA, Anderson FA Jr, Dabbous OH et al. Intervention in acute coronary syndromes: do patients undergo intervention on the basis of their risk characteristics? The Global Registry of Acute Coronary Events (GRACE). *Heart*. 2007;93:177-182.
 22. Bebb O, Smith FG, Clegg A, Hall M, Gale CP. Frailty and acute coronary syndrome: A structured literature review. *Eur Heart J Acute Cardiovasc Care*. 2018;7:166-175.
 23. Nicholson G, Gandra SR, Halbert RJ, Richhariya A, Nordyke RJ. Patient-level costs of major cardiovascular conditions: a review of the international literature. *Clinicoecon Outcomes Res*. 2016;8:495-506.
 24. Demaerschalk BM, Hwang HM, Leung G. US cost burden of ischemic stroke: a systematic literature review. *Am J Manag Care*. 2010;16:525-533.
 25. Luengo-Fernandez R, Gray AM, Rothwell PM. Costs of stroke using patient-level data: a critical review of the literature. *Stroke*. 2009;40:e18-e23.
 26. Payne KA, Huybrechts KF, Caro JJ, Craig Green TJ, Klittich WS. Long term cost-of-illness in stroke: an international review. *Pharmacoeconomics*. 2002;20:813-825.
 27. Reynolds MW, Frame D, Scheye R et al. A systematic review of the economic burden of chronic angina. *Am J Manag Care*. 2004;10(11 suppl):S347-S357.
 28. García-Blas S, Bonanad C, Sanchis J. Invasive strategy in elderly patients with acute coronary syndrome in 2018: close to the truth? *J Geriatr Cardiol*. 2019;16:114-120.

29. André R, Bongard V, Elosua R et al. International differences in acute coronary syndrome patients' baseline characteristics, clinical management and outcomes in Western Europe: the EURHOBOP study. *Heart*. 2014;100:1201-1207.

Supplementary material

Tables

Supplementary Table 1. Demographic and clinical characteristics of the study populations

Panel A	EUROTRACS cohort (n=28,600)
Age	82.1 (5.16)
Sex: Female	14,014 (49.0%)
Diabetes	6,034 (21.1%)
Hypertension	11,032 (38.6%)
Chronic kidney disease	3,414 (11.9%)
Previous MI	5,948 (20.8%)
Admission Killip class: III-IV	1,931 (6.75%)
Type of ACS	
Non-ST-segment elevation ACS	16,621 (58.1%)
ST-segment elevation MI	9,975 (34.9%)
Non classifiable	2,004 (7.01%)
Admission year*	
2000-2004	4,590 (16.2%)
2005-2009	7,272 (25.6%)
2010-2014	16,498 (58.2%)
Thrombolysis	1,036 (3.62%)
PCI	8,897 (31.1%)
CABG	426 (1.49%)
Maximum Killip class: III-IV	1,975 (6.91%)
In-hospital mortality	3,957 (13.8%)
Panel B	SIDIAP cohort (n=55,564)
Age	68.9 (13.0)
Sex: Female	17,827 (32.1%)
Diabetes	18,334 (33.0%)
Hypertension	38,702 (69.7%)
Chronic kidney disease	5,621 (10.1%)
Type of ACS	
AMI	52,371 (94.3%)
UA	3,193 (5.75%)
Admission year	
2007-2009	26,600 (47.9%)
2010-2013	28,964 (52.1%)
PCI	1,478 (2.66%)
CABG	18,165 (32.7%)
Recurrence	15,851 (28.5%)
Death (with/without prior recurrence)	3,303 (5.94%) / 7,880 (14.2%)

* Admission date was missing for 240 ACS patients. ACS, acute coronary syndrome; CABG, coronary artery bypass grafting; MI, myocardial infarction; PCI, percutaneous coronary intervention.

Supplementary Table 2. PCI use and in-hospital mortality rates according to sex and PS tertile

Sex	PS tertile	Basal rate of use of PCI	Mortality rate		Mortality reduction
			Without PCI	With PCI	
Females	1 st	13.1%	19.2%	10.8%	8.4%
	2 nd	25.1%	13.8%	7.6%	6.2%
	3 rd	40.9%	19.4%	11.0%	8.4%
Males	1 st	20.3%	18.9%	8.5%	10.4%
	2 nd	33.7%	12.7%	7.5%	5.2%
	3 rd	49.7%	15.2%	7.7%	7.5%

Tertile 1: lowest probability of receiving PCI; Tertile 2: intermediate probability of receiving PCI; Tertile 3: highest probability of receiving PCI. PCI, percutaneous coronary intervention; PS, propensity score.

Supplementary Table 3. Multistate transition model

Transition	Factor	HR (95% CI)	p-value
From primary ACS to recurrence	Age group (ref 30-49)		
	50-59	1.18 (1.10 – 1.26)	<0.001
	60-69	1.28 (1.20 – 1.36)	<0.001
	70-79	1.42 (1.34 – 1.51)	<0.001
	80-89	1.32 (1.23 – 1.41)	<0.001
	90+	1.13 (0.99 – 1.29)	0.073
	Male (ref female)	1.20 (1.16 – 1.25)	<0.001
	AMI (ref UA)	0.65 (0.60 – 0.71)	<0.001
	PCI yes (ref no)	0.91 (0.88 – 0.95)	<0.001
From primary ACS to death	Age group (ref 30-49)		
	50-59	1.62 (1.30 – 2.02)	<0.001
	60-69	2.91 (2.38 – 3.57)	<0.001
	70-79	7.20 (5.92 – 8.75)	<0.001
	80-89	18.97 (15.62 – 23.05)	<0.001
	90+	48.53 (39.56 – 59.53)	<0.001
	Male (ref female)	1.23 (1.18 – 1.29)	<0.001
	AMI (ref UA)	1.16 (1.07 – 1.25)	<0.001
	PCI yes (ref no)	0.51 (0.48 – 0.55)	<0.001
From recurrence to death	Age group (ref 30-49)		
	50-59	1.77 (1.26 – 2.47)	<0.001
	60-69	3.50 (2.56 – 4.78)	<0.001
	70-79	7.00 (5.17 – 9.50)	<0.001
	80-89	18.07 (13.32 – 24.51)	<0.001
	90+	41.40 (29.67 – 57.78)	<0.001
	Male (ref female)	1.13 (1.05 – 1.22)	<0.001
	AMI (ref UA)	1.55 (1.34 – 1.81)	<0.001
	PCI yes (ref no)	0.51 (0.45 – 0.57)	<0.001

ACS, acute coronary syndrome; AMI, acute myocardial infarction; CI, confidence interval; HR, hazard ratio; PCI, percutaneous coronary intervention; UA, unstable angina.

Supplementary Table 4. Unit costs by country (2018 euros)

Cost variables	Unit	France	Germany	Greece	Italy	Portugal	Spain
Cost of PCI	1 intervention	3,147.64	3,139.06	2,023.48	3,080.73	2,223.10	3,607.68
Acute care costs							
Surviving AMI patients with complications	1 event	3,559.10	8,469.82	1,837.50	4,823.26	3,906.30	5,216.22
Surviving AMI patients without complications	1 event	1,421.98	3,732.99	729.75	3,465.56	2,461.09	4,893.98
AMI patients who died	1 event	2,777.63	11,021.22	1,837.50	4,123.37	3,261.80	3,511.44
Surviving UA patients with complications	1 event	3,069.91	7,633.29	950.08	1,919.04	3,906.30	5,216.22
Surviving UA patients without complications	1 event	865.42	2,258.62	428.55	1,919.04	2,461.09	4,893.98
UA patients who died	1 event	852.98	14,409.14	950.08	1,919.04	3,261.80	3,511.44
Follow-up costs							
AMI	1 year	886.56	771.34	1,678.50	3,369.99	446.15	1,430.51
UA	1 year	1,066.07	968.21	1,283.95	2,040.11	656.98	1,203.07

AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; UA, unstable angina.

Supplementary Table 5. Values for deterministic one-way sensitivity analysis

Model parameter	Left values	Right values	Base case values
Female			
In-hospital mortality reduction (PCI vs non-PCI)			
<i>1st tertile of PS</i>	6.30%	10.50%	8.40%
<i>2nd tertile of PS</i>	4.65%	7.75%	6.20%
<i>3rd tertile of PS</i>	6.30%	10.50%	8.40%
Male			
In-hospital mortality reduction (PCI vs non-PCI)			
<i>1st tertile of PS</i>	7.80%	13.00%	10.40%
<i>2nd tertile of PS</i>	3.90%	6.50%	5.20%
<i>3rd tertile of PS</i>	5.63%	9.38%	7.50%
Common parameters			
Acute care costs*			
<i>Surviving AMI patients with complications</i>	2,669.33	4,448.88	3,559.10
<i>Surviving AMI patients without complications</i>	1,066.49	1,777.48	1,421.98
<i>AMI patients who died</i>	2,083.22	3,472.04	2,777.63
<i>Surviving UA patients with complications</i>	2,302.43	3,837.38	3,069.91
<i>Surviving UA patients without complications</i>	649.06	1,081.77	865.42
<i>UA patients who died</i>	639.74	1,066.23	852.98
<i>PCI</i>	2,360.73	3,934.55	3,147.64
Follow-up costs per year*			
<i>UA</i>	799.55	1,332.59	1,066.07
<i>AMI</i>	664.92	1,108.20	886.56
Patients hospitalized with complications	34.5%	57.4%	45.9%
Disutilities			
<i>ACS first year</i>	0.0791	0.1319	0.1055
<i>ACS subsequent years</i>	0.0750	0.1250	0.1000
<i>Death</i>	0.7500	1.0000	1.0000
Annual discount rate	0.00%	5.00%	3.00%

* Cost values ranged between 0.75 and 1.25 times its baseline value. Values shown correspond to France.

ACS, acute coronary syndrome; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; PS, propensity score; UA, unstable angina.

Supplementary Table 6. Distributions assumed for probabilistic sensitivity analysis

Model parameter	Distribution	Alpha	Beta
Female			
In-hospital mortality reduction (PCI vs non-PCI)			
<i>1st tertile of PS</i>	Beta	1,861	20,294
<i>2nd tertile of PS</i>	Beta	1,037	15,690
<i>3rd tertile of PS</i>	Beta	909	9,907
Male			
In-hospital mortality reduction (PCI vs non-PCI)			
<i>1st tertile of PS</i>	Beta	2,304	19,851
<i>2nd tertile of PS</i>	Beta	870	15,857
<i>3rd tertile of PS</i>	Beta	811	10,005
Common parameters			
Acute care costs*			
<i>Surviving AMI patients with complications</i>	Gamma	100	35.59
<i>Surviving AMI patients without complications</i>	Gamma	100	14.22
<i>AMI patients who died</i>	Gamma	100	27.78
<i>Surviving UA patients with complications</i>	Gamma	100	30.70
<i>Surviving UA patients without complications</i>	Gamma	100	8.65
<i>UA patients who died</i>	Gamma	100	8.53
<i>PCI</i>	Gamma	100	31.48
Follow-up costs per year*			
<i>UA</i>	Gamma	100	10.66
<i>AMI</i>	Gamma	100	8.87
Patients hospitalized with complications	Beta	39,576	46,566
Disutilities			
<i>ACS first year</i>	Lognormal	-2.25	0.011
<i>ACS subsequent years</i>	Lognormal	-2.30	0.010
<i>Death</i>	Lognormal	0.00	0.150

* Coefficient of variation for the cost parameters (the ratio of the standard error to the mean) was assumed to be 0.1. Values shown correspond to France.

ACS, acute coronary syndrome; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; PS, propensity score; UA, unstable angina.

Supplementary Table 7. Effect of PCI on in-hospital mortality in elderly patients (≥ 75 years)

Tertile of PS	Female OR (95% CI)	Male OR (95% CI)
1 st	0.47 (0.34, 0.64)	0.34 (0.26, 0.46)
2 nd	0.43 (0.33, 0.57)	0.48 (0.37, 0.62)
3 rd	0.43 (0.35, 0.53)	0.40 (0.32, 0.50)

Tertile 1: lowest probability of receiving PCI; Tertile 2: intermediate probability of receiving PCI; Tertile 3: highest probability of receiving PCI. CI, confidence interval; OR, odds-ratio; PCI, percutaneous coronary intervention; PS, propensity score.

Supplementary Table 8. Results of the deterministic one-way sensitivity analysis for France by sex

Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	4.3	4.4	4.6	4.6	4.5	4.5	4.3	4.4
PCI vs non-PCI fatality rate reduction	4.9	3.9	5.2	4.2	5.1	4.0	4.9	3.9
Disutilities	5.8	4.5	6.1	4.7	6.0	4.6	5.8	4.5
Acute care costs (AMI)	4.4	4.3	4.7	4.5	4.5	4.4	4.4	4.3
Acute care costs (UA)	4.3	4.3	4.6	4.6	4.5	4.5	4.3	4.3
Cost of PCI	3.4	5.3	3.6	5.6	3.5	5.5	3.4	5.3
Follow-up costs (AMI)	4.3	4.4	4.5	4.6	4.4	4.5	4.3	4.4
Follow-up costs (UA)	4.2	4.5	4.4	4.8	4.3	4.7	4.2	4.5
Annual discount rate	3.9	4.6	4.1	4.9	4.1	4.8	3.9	4.6
	Male							
Patients hospitalized with complications	3.9	4.0	4.2	4.2	3.9	3.9	3.5	3.6
PCI vs non-PCI fatality rate reduction	4.4	3.6	4.6	3.9	4.3	3.6	4.0	3.2
Disutilities	5.3	4.1	5.6	4.3	5.2	4.0	4.7	3.6
Acute care costs (AMI)	4.0	3.9	4.3	4.1	4.0	3.8	3.6	3.5
Acute care costs (UA)	3.9	4.0	4.2	4.2	3.9	3.9	3.5	3.5
Cost of PCI	3.1	4.8	3.3	5.1	3.0	4.7	2.8	4.3
Follow-up costs (AMI)	3.9	4.0	4.2	4.2	3.9	3.9	3.5	3.6
Follow-up costs (UA)	3.8	4.1	4.0	4.3	3.7	4.0	3.4	3.7
Annual discount rate	3.6	4.2	3.8	4.5	3.5	4.1	3.2	3.8

ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

Supplementary Table 9. Results of the deterministic one-way sensitivity analysis for Germany by sex

Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	2.7	2.8	2.8	2.9	2.8	2.9	2.7	2.8
PCI vs non-PCI fatality rate reduction	3.2	2.4	3.3	2.6	3.2	2.5	3.2	2.4
Disutilities	3.7	2.8	3.9	3.0	3.8	2.9	3.7	2.8
Acute care costs (AMI)	3.1	2.4	3.2	2.5	3.2	2.5	3.1	2.4
Acute care costs (UA)	2.9	2.7	3.0	2.8	2.9	2.7	2.8	2.7
Cost of PCI	1.8	3.7	1.9	3.9	1.9	3.8	1.8	3.7
Follow-up costs (AMI)	2.7	2.8	2.9	2.9	2.8	2.9	2.7	2.8
Follow-up costs (UA)	2.6	2.9	2.7	3.0	2.7	3.0	2.6	2.9
Annual discount rate	2.4	3.0	2.5	3.1	2.5	3.1	2.4	3.0
	Male							
Patients hospitalized with complications	2.2	2.4	2.4	2.5	2.2	2.3	2.0	2.2
PCI vs non-PCI fatality rate reduction	2.6	2.1	2.7	2.2	2.5	2.1	2.4	1.9
Disutilities	3.1	2.4	3.2	2.5	3.0	2.3	2.8	2.1
Acute care costs (AMI)	2.6	1.9	2.8	2.0	2.6	1.9	2.4	1.8
Acute care costs (UA)	2.4	2.2	2.5	2.3	2.4	2.2	2.2	2.0
Cost of PCI	1.4	3.2	1.5	3.4	1.4	3.1	1.3	2.8
Follow-up costs (AMI)	2.3	2.3	2.4	2.5	2.2	2.3	2.0	2.1
Follow-up costs (UA)	2.2	2.4	2.3	2.6	2.1	2.4	1.9	2.2
Annual discount rate	2.0	2.5	2.1	2.6	2.0	2.5	1.8	2.2

ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

Supplementary Table 10. Results of the deterministic one-way sensitivity analysis for Greece by sex

Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	3.6	3.7	3.8	3.8	3.7	3.8	3.6	3.7
PCI vs non-PCI fatality rate reduction	4.1	3.4	4.2	3.5	4.1	3.5	4.0	3.4
Disutilities	4.9	3.8	5.1	3.9	5.0	3.9	4.9	3.8
Acute care costs (AMI)	3.7	3.6	3.9	3.8	3.8	3.7	3.7	3.6
Acute care costs (UA)	3.7	3.7	3.8	3.8	3.8	3.8	3.7	3.7
Cost of PCI	3.1	4.3	3.2	4.5	3.1	4.4	3.1	4.3
Follow-up costs (AMI)	3.6	3.7	3.8	3.9	3.7	3.8	3.6	3.7
Follow-up costs (UA)	3.3	4.0	3.5	4.2	3.4	4.1	3.3	4.0
Annual discount rate	3.4	3.8	3.5	4.0	3.5	3.9	3.4	3.8
	Male							
Patients hospitalized with complications	3.3	3.3	3.5	3.5	3.3	3.3	3.0	3.1
PCI vs non-PCI fatality rate reduction	3.6	3.1	3.8	3.3	3.6	3.1	3.3	2.8
Disutilities	4.4	3.4	4.7	3.6	4.4	3.4	4.1	3.1
Acute care costs (AMI)	3.4	3.3	3.5	3.4	3.3	3.2	3.1	3.0
Acute care costs (UA)	3.3	3.3	3.5	3.5	3.3	3.3	3.1	3.0
Cost of PCI	2.8	3.9	2.9	4.1	2.7	3.8	2.6	3.5
Follow-up costs (AMI)	3.3	3.4	3.4	3.5	3.2	3.3	3.0	3.1
Follow-up costs (UA)	3.0	3.6	3.2	3.8	3.0	3.6	2.8	3.3
Annual discount rate	3.1	3.5	3.2	3.7	3.0	3.5	2.8	3.2

ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

Supplementary Table 11. Results of the deterministic one-way sensitivity analysis for Italy by sex

Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	6.2	6.2	6.4	6.4	6.3	6.3	6.2	6.2
PCI vs non-PCI fatality rate reduction	6.7	5.8	7.0	6.0	6.9	5.9	6.7	5.8
Disutilities	8.3	6.4	8.6	6.6	8.4	6.5	8.2	6.4
Acute care costs (AMI)	6.3	6.1	6.5	6.3	6.4	6.2	6.3	6.1
Acute care costs (UA)	6.2	6.2	6.4	6.4	6.3	6.3	6.2	6.2
Cost of PCI	5.3	7.1	5.4	7.4	5.4	7.3	5.3	7.1
Follow-up costs (AMI)	6.1	6.3	6.3	6.5	6.3	6.4	6.1	6.3
Follow-up costs (UA)	5.5	6.8	5.8	7.1	5.7	7.0	5.5	6.8
Annual discount rate	5.8	6.5	6.0	6.7	5.9	6.6	5.8	6.5
	Male							
Patients hospitalized with complications	5.6	5.6	5.8	5.8	5.5	5.6	5.2	5.3
PCI vs non-PCI fatality rate reduction	6.0	5.3	6.2	5.5	5.9	5.3	5.6	4.9
Disutilities	7.5	5.8	7.8	6.0	7.4	5.7	7.0	5.4
Acute care costs (AMI)	5.7	5.5	5.9	5.7	5.6	5.5	5.3	5.2
Acute care costs (UA)	5.6	5.6	5.8	5.8	5.6	5.5	5.2	5.2
Cost of PCI	4.8	6.5	4.9	6.7	4.7	6.4	4.5	6.0
Follow-up costs (AMI)	5.5	5.7	5.8	5.9	5.5	5.6	5.2	5.3
Follow-up costs (UA)	5.0	6.2	5.2	6.4	5.0	6.1	4.7	5.8
Annual discount rate	5.2	5.9	5.4	6.1	5.2	5.8	4.9	5.5

ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

Supplementary Table 12. Results of the deterministic one-way sensitivity analysis for Portugal by sex

Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	2.8	2.8	2.9	2.9	2.8	2.9	2.8	2.8
PCI vs non-PCI fatality rate reduction	3.1	2.5	3.3	2.6	3.2	2.6	3.1	2.5
Disutilities	3.7	2.9	3.9	3.0	3.8	2.9	3.7	2.9
Acute care costs (AMI)	2.8	2.7	3.0	2.8	2.9	2.8	2.8	2.7
Acute care costs (UA)	2.8	2.8	2.9	2.9	2.9	2.8	2.8	2.8
Cost of PCI	2.1	3.4	2.2	3.6	2.2	3.5	2.1	3.4
Follow-up costs (AMI)	2.8	2.8	2.9	2.9	2.8	2.9	2.7	2.8
Follow-up costs (UA)	2.7	2.9	2.8	3.0	2.8	2.9	2.7	2.9
Annual discount rate	2.5	3.0	2.6	3.1	2.6	3.1	2.5	3.0
	Male							
Patients hospitalized with complications	2.5	2.5	2.6	2.6	2.4	2.5	2.2	2.3
PCI vs non-PCI fatality rate reduction	2.7	2.3	2.9	2.4	2.7	2.3	2.5	2.1
Disutilities	3.3	2.6	3.5	2.7	3.3	2.5	3.0	2.3
Acute care costs (AMI)	2.6	2.4	2.7	2.5	2.5	2.4	2.3	2.2
Acute care costs (UA)	2.5	2.5	2.6	2.6	2.5	2.4	2.3	2.3
Cost of PCI	1.9	3.1	2.0	3.3	1.9	3.1	1.7	2.8
Follow-up costs (AMI)	2.5	2.5	2.6	2.7	2.4	2.5	2.2	2.3
Follow-up costs (UA)	2.4	2.6	2.6	2.7	2.4	2.5	2.2	2.3
Annual discount rate	2.2	2.7	2.3	2.8	2.2	2.6	2.0	2.4

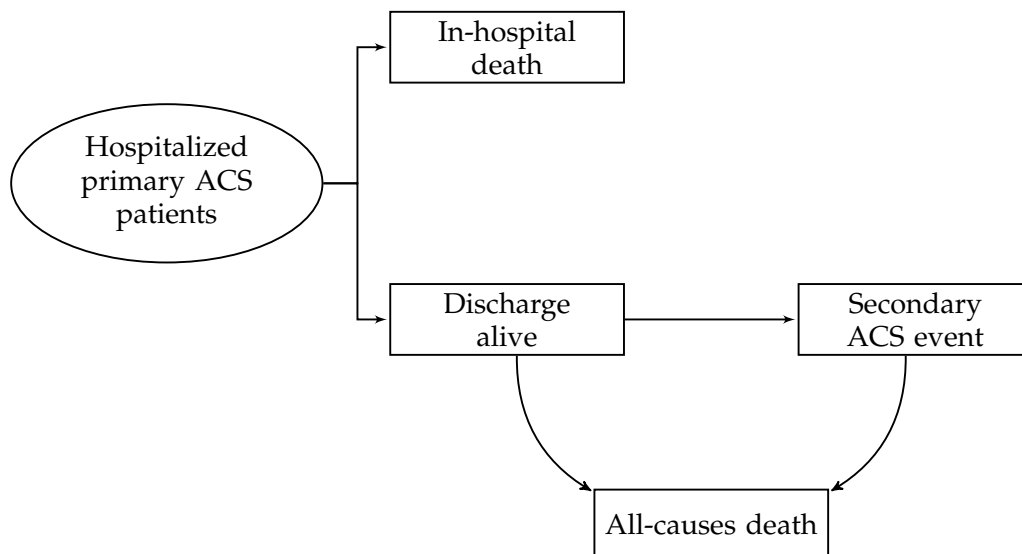
ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

Supplementary Table 13. Results of the deterministic one-way sensitivity analysis for Spain by sex

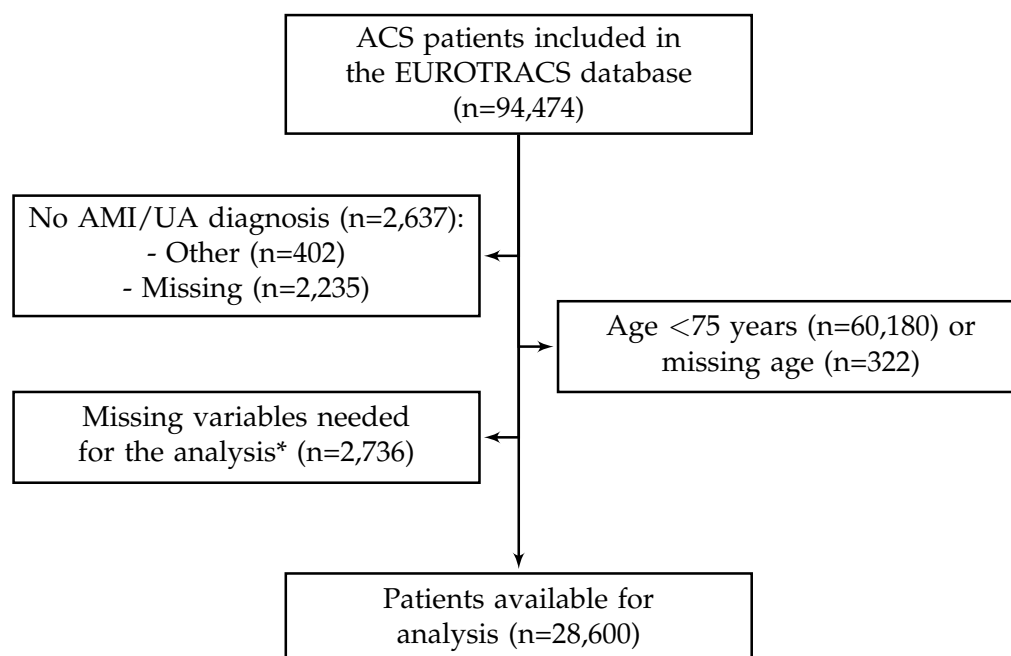
Parameter	Scenario A		Scenario B		Scenario C		Scenario D	
	Left	Right	Left	Right	Left	Right	Left	Right
	Female							
Patients hospitalized with complications	5.4	5.4	5.7	5.7	5.6	5.6	5.4	5.4
PCI vs non-PCI fatality rate reduction	6.0	5.0	6.3	5.2	6.2	5.1	6.0	5.0
Disutilities	7.2	5.6	7.6	5.8	7.4	5.7	7.2	5.6
Acute care costs (AMI)	5.5	5.4	5.7	5.6	5.6	5.5	5.4	5.4
Acute care costs (UA)	5.4	5.4	5.7	5.7	5.6	5.6	5.4	5.4
Cost of PCI	4.4	6.5	4.5	6.8	4.4	6.7	4.3	6.5
Follow-up costs (AMI)	5.4	5.5	5.6	5.7	5.5	5.6	5.4	5.5
Follow-up costs (UA)	5.1	5.7	5.4	5.9	5.3	5.8	5.1	5.7
Annual discount rate	4.9	5.8	5.1	6.0	5.1	5.9	4.9	5.8
	Male							
Patients hospitalized with complications	4.9	4.9	5.2	5.2	4.9	4.9	4.5	4.5
PCI vs non-PCI fatality rate reduction	5.4	4.6	5.6	4.8	5.3	4.5	5.0	4.2
Disutilities	6.6	5.1	6.9	5.3	6.5	5.0	6.0	4.7
Acute care costs (AMI)	5.0	4.9	5.2	5.1	4.9	4.8	4.5	4.5
Acute care costs (UA)	4.9	4.9	5.2	5.2	4.9	4.9	4.5	4.5
Cost of PCI	3.9	5.9	4.1	6.2	3.9	5.8	3.7	5.4
Follow-up costs (AMI)	4.9	5.0	5.1	5.2	4.8	4.9	4.5	4.6
Follow-up costs (UA)	4.7	5.2	4.9	5.4	4.6	5.1	4.3	4.8
Annual discount rate	4.5	5.2	4.7	5.5	4.4	5.2	4.1	4.8

ICUR values expressed as incremental cost ($\times 10^6$ euros) per quality-adjusted life year gained. AMI, acute myocardial infarction; ICUR, incremental cost-utility ratio; UA, unstable angina; PCI, percutaneous coronary intervention.

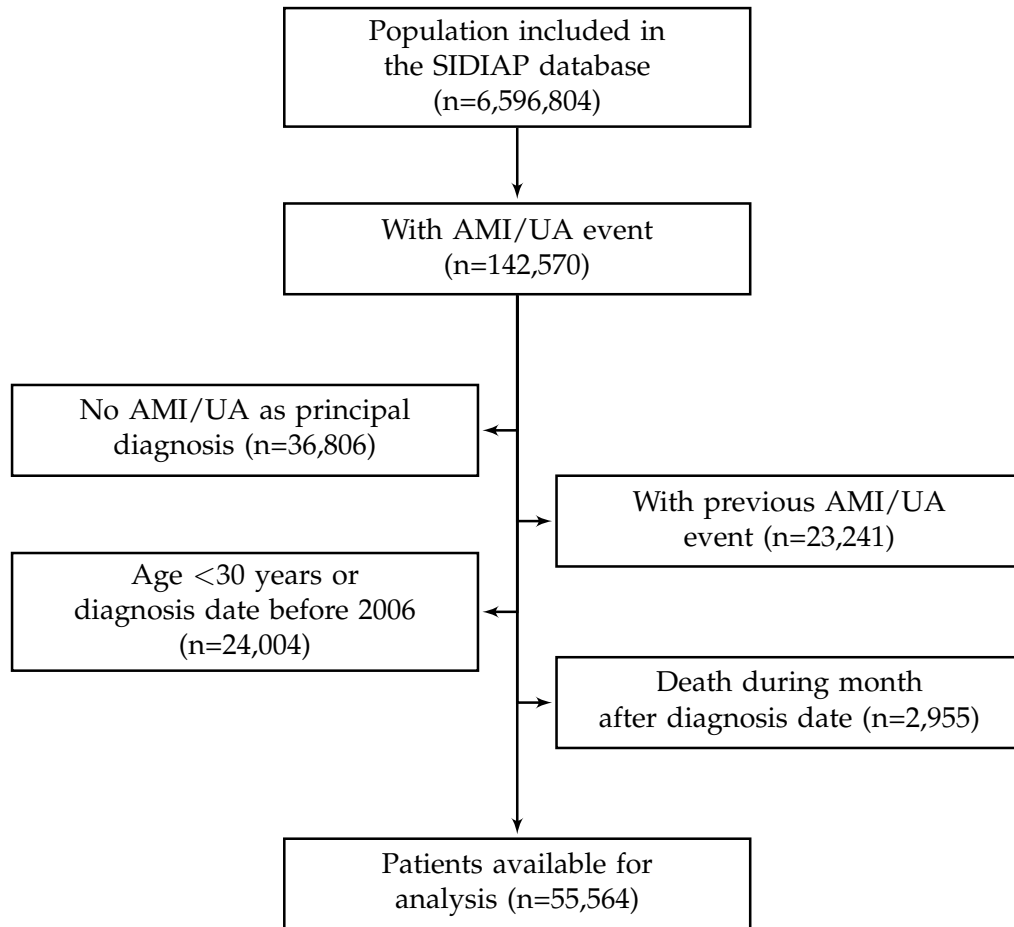
Figures



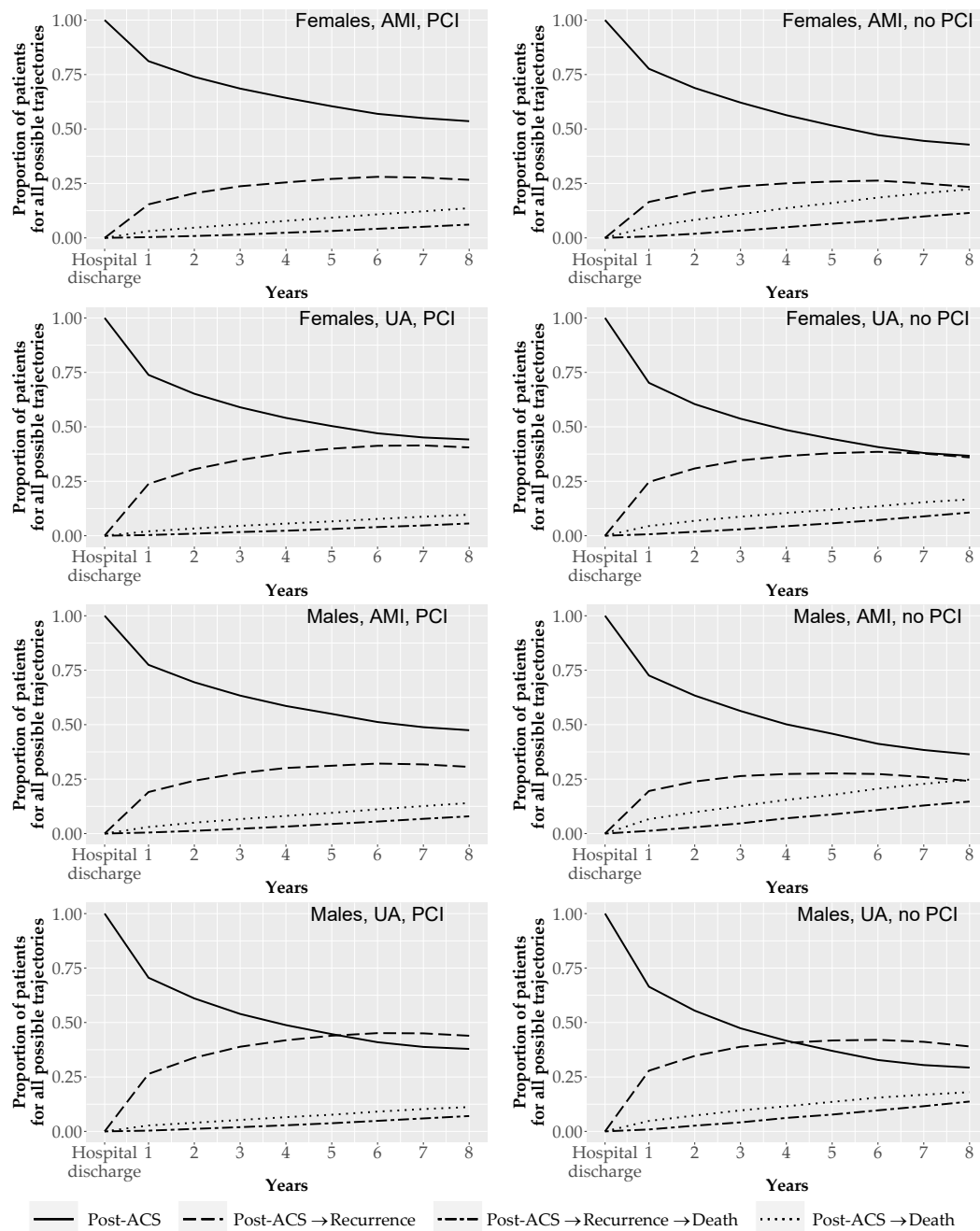
Supplementary Figure 1. Structure of the semi-Markov state transition model. Semi-Markov process with identical health states for each of the management options. In-hospital mortality rates were obtained from the EUROTRACS ACS database. All patients discharged alive remained in this state until recurrence or death occurs. The transition probabilities were obtained from the SIDIAP database.



Supplementary Figure 2. Flowchart of the participants included from the EUROTRACS ACS database. * Variables needed for the analysis: ACS type, admission Killip, age, diabetes, hypertension, maximum Killip, PCI use, previous MI, thrombolysis use. ACS, acute coronary syndrome; AMI, acute myocardial infarction; UA, unstable angina.



Supplementary Figure 3. Flowchart of the population included from the SIDIAP database. AMI, acute myocardial infarction; UA, unstable angina.



Supplementary Figure 4. Proportion of patients for all possible trajectories after hospital discharge alive by sex, primary ACS event and PCI. Predicted probabilities calculated by sampling all possible trajectories through the multistate model implemented with SIDIAP data (10,000 bootstrap data sets). ACS, acute coronary syndrome; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; UA, unstable angina.

Capítol 4

Discussió

4.1 Resultats principals

Els resultats dels manuscrits inclosos en la tesi sostenen les hipòtesis principals. Per una banda, indiquen la idoneïtat d'adaptar els programes de cribratge de càncer de mama al risc de les dones, ja que d'aquesta manera s'obtidrien beneficis en salut similars, reduint costos i efectes adversos, o proporcionarien millors resultats en salut amb el mateix pressupost. També mostren que amb la incorporació de mesures longitudinals de la densitat mamària als models de risc de càncer de mama, junt amb els altres factors de risc coneguts, es podrien obtenir estimacions de risc més acurades.

Per una altra banda, els costos addicionals associats a l'increment de l'ús de la intervenció coronària percutània (ICP) en població anciana ingressada per un esdeveniment de síndrome coronària aguda (SCA) serien sobradament compensats pel major benefici en salut obtingut per la reducció de la mortalitat hospitalària i el millor pronòstic després de l'alta hospitalària associats a la ICP. Finalment, també s'han identificat els biomarcadors de malaltia cardiovascular en pacients amb malaltia renal crònica (MRC) més prometedors per continuar investigant-ne la incorporació en models d'estimació de risc cardiovascular. A continuació s'enumeren els resultats principals d'aquesta tesi.

Sobre les estratègies de cribratge poblacional de càncer de mama i l'estimació del risc de càncer de mama:

1. En comparació amb les estratègies uniformes, les basades en el risc produeixen un benefici molt superior per a un pressupost determinat. Les estratègies basades en el risc proporcionen reduccions properes al 10% en els costos i superiors al 20% en resultats de falsos positius i casos sobrediagnosticats.
2. El cribratge òptim es caracteritza per periodicitats quinquennals o triennals per als grups de risc baix o moderat i una periodicitat anual per al grup d'alt risc.

3. Un augment de la densitat mamària amb l'edat s'associa a un augment del risc de càncer de mama. Alternativament, una disminució de la densitat s'associa a una reducció del risc.
4. La modelització conjunta de la història longitudinal completa de densitat mamària i de l'edat al diagnòstic de càncer de mama permet estimar el risc de càncer de mama individual de manera dinàmica, així com realitzar prediccions de valors futurs del marcador longitudinal.

Pel que fa a la millora de l'estimació del risc cardiovascular en pacients amb MRC:

5. Tot i que no s'ha trobat cap millora rellevant en la capacitat de discriminació afegida pels potencials biomarcadors als factors de risc ja coneguts, l'osteopontina va ser el biomarcador amb major capacitat predictiva d'esdeveniment cardiovascular, per davant de factors de risc ben descrits en la literatura. L'osteoprotegerina, la metal·loproteïnasa de matriu 9 i el factor de creixement de l'endoteli vascular també van destacar per la capacitat de discriminació.

Finalment, sobre l'avaluació de l'increment de l'ús de la ICP en pacients ancians ingressats per SCA:

6. L'avaluació econòmica mostra que augmentar el nombre de pacients ancians amb SCA que reben una ICP és rendible en un llinar de disposició a pagar de 10.000 € per any de vida ajustat per qualitat (VAQ) guanyat, en un horitzó temporal de 8 anys.
7. Tots els escenaris alternatius avaluats van proporcionar resultats d'eficiència molt similars, els quals reflecteixen la reducció de mortalitat hospitalària per ús de la ICP observada en l'estudi EUROTRACS. Els resultats també suggereixen que l'impacte de l'ús de la ICP en els resultats de salut és més gran en el període hospitalari que en els 8 anys següents.

4.2 Discussió conjunta

Els models matemàtics i estadístics utilitzats en els treballs inclosos en aquesta tesi han permès aportar evidència rellevant en les seves respectives àrees d'interès i han proporcionat informació útil per a la presa de decisions. Així, utilitzant models probabilístics simulant la història natural del càncer de mama, s'ha mostrat que les estratègies de cribratge basades en el risc són més eficients i produeixen relacions de cost-efectivitat i de dany-benefici més baixes que les estratègies uniformes. Per exemple, si en lloc d'oferir un programa de cribratge biennal a totes les dones de 50 a 69 anys, es combinessin periodicitats d'exàmens quinquennals, triennals i anuals per a dones amb risc baix o moderat-baix, moderat-alt i alt, respectivament, en l'interval d'edat de 50 a 74 anys, s'evitara el mateix nombre de morts. De manera similar, les estratègies que combinen exàmens quinquennals per a dones amb risc baix o moderat-baix amb exàmens anuals per a dones amb risc moderat-alt

o alt, en l'interval d'edat de 45 a 74 anys, suposen un benefici similar en AVAQ que l'estratègia biennal uniforme en el mateix interval d'edat. Però el resultat important és que en ambdós casos les estratègies basades en el risc comportarien reduccions notables de costos, resultats de falsos positius de la mamografia i casos sobrediagnosticats, tant de càncer invasiu com *in situ*. Diversos treballs basats en models de decisió han proposat recomanacions personalitzades per al cribratge de càncer de mama basades en anàlisis de cost-efectivitat, de cost-utilitat o comparant-ne els danys i els beneficis [157-164]. Amb les diferències metodològiques existents entre si, tots conclouen que el cribratge basat en el risc millora els beneficis en salut i redueix els danys i els costos d'un cribratge uniforme independent del risc. Per exemple, Schousboe *et al.* [157] van establir uns llindars de cost-utilitat de 100.000 \$ o 50.000 \$ per AVAQ i van comparar diferents periodicitats per als exàmens de cribratge en grups d'edat de deu anys, categories de densitat mamària en l'escala BI-RADS i la presència o absència d'història personal de biòpsia i història familiar de càncer de mama. Van recomanar que les dones de 50 a 79 anys amb una baixa densitat mamària i sense cap factor de risc de càncer de mama podien considerar fer-se mamografies amb menys freqüència que cada 2 anys, resultat consistent amb els nostres resultats. Però van recomanar un cribratge biennal per a dones de 50 a 79 anys amb densitat mamària alta —categories BI-RADS *c* i *d*—, independentment de la presència d'altres factors de risc. En el nostre estudi, les dones en les categories *c* o *d* de densitat mamària pertanyen a grups de risc diferents, dependent també de la presència de 0, 1 o 2 factors de risc addicionals; per tant, l'estratègia òptima recomanaria diferents periodicitats i intervals d'edat per als diferents grups de risc. A més a més, Schousboe *et al.* van concloure que la mamografia anual no era eficient per a cap grup, independentment de l'edat o de la densitat mamària, fet que tampoc està en línia amb els nostres resultats, probablement degut a les diferències en els objectius dels estudis i en els enfocaments metodològics utilitzats.

El nostre treball no pretén ser una guia detallada de com s'hauria de fer el cribratge personalitzat, sinó una avaluació global de l'impacte que un nou paradigma de cribratge tindria sobre els costos i els resultats en salut, tant beneficis com danys. Una revisió sistemàtica duta a terme per un investigador independent va avaluar els models de simulació de cribratge estratificat pel risc de càncer de mama [165]. Entre tots els treballs inclosos en la revisió, el nostre estudi va ser avaluat com el millor en la simulació del cribratge en població general, per davant de l'excel·lent treball publicat per Schousboe *et al.* uns anys abans.

Per tal d'adaptar el cribratge al risc individual de càncer de mama i avançar en l'atenció sanitària personalitzada cal disposar de mesures acurades del risc. Hi ha diversos instruments dissenyats amb aquesta finalitat, tot i que inicialment van ser ideats per identificar dones que poguessin beneficiar-se de quimioprevenió primària. Probablement, el més conegut de tots és el model de Gail [166], que ha servit de referència per al desenvolupament d'altres models. Chen *et al.* [167] i Barlow *et al.* [168] van desenvolupar

nous models que, entre altres factors de risc, incorporaven la densitat mamària, un dels factors que més explica la variabilitat de la incidència de càncer de mama. Diversos estudis han mostrat que una densitat mamària elevada s'associa amb un major risc de càncer de mama, amb estimacions de risc 4-6 vegades major entre els quintils extrems de la població [169, 170]. Les lesions benignes prèvies i les variants gèniques o SNP (de l'anglès, *single-nucleotide polymorphism*) han estat les variables més rellevants incorporades als models de predicció de risc de càncer de mama en els darrers anys [171-174]. Amb tot, tal com s'ha vist en una revisió sistemàtica recent, la capacitat de discriminació dels models continua sent limitada [175]. En el segon treball d'aquesta tesi, s'ha proposat un model conjunt bayesià que combina la història longitudinal completa de la densitat mamària i l'edat en el diagnòstic de càncer de mama, juntament amb dos dels factors de risc de càncer de mama més coneguts —antecedents familiars de càncer de mama i procediments previs de mama. El model conjunt, estimat amb les dades d'una cohort de dones participants en un programa poblacional de cribratge de càncer de mama, va proporcionar resultats coherents amb la literatura sobre l'impacte de l'evolució de la densitat mamària en el risc de càncer de mama [176-179], i va mostrar un increment de risc de diagnòstic de càncer de mama en aquelles dones que no experimentaven una disminució de la densitat mamària amb l'edat. Considerem que el nostre model proporciona un bon punt de partida, a partir del qual en caldria estendre i avaluar el calibratge i la discriminació, per intentar augmentar la capacitat predictiva dels models actuals.

En el tercer treball d'aquesta tesi s'ha explorat, mitjançant algorismes d'aprenentatge automàtic, l'aportació de dinou biomarcadors sèrics en la millora de la predicció del risc cardiovascular en una cohort composta principalment per pacients amb MRC. Si bé l'addició dels nous biomarcadors potencials va millorar marginalment la capacitat de discriminació obtinguda amb només els factors de risc cardiovascular coneguts en MRC, l'osteopontina va ser el predictor més rellevant en l'anàlisi per models de bosc aleatori de supervivència, fins i tot per davant de factors de risc ben descrits com el sexe masculí, l'augment de l'edat, el tabaquisme, la diabetis mellitus i el colesterol total elevat [180]. Aquest resultat va ser confirmat en l'anàlisi de regressió de Fine i Gray, en la qual nivells alts d'osteopontina i de tres biomarcadors més —osteoprotegerina, metal·loproteïnasa de matriu 9, i factor de creixement de l'endoteli vascular— també van mostrar un increment significatiu del risc cardiovascular. Els resultats són consistents en relació amb estudis previs en què s'avaluava l'impacte en el risc cardiovascular de l'osteopontina [181] i l'osteoprotegerina [182], o la relació amb processos inflamatoris i la formació i desestabilització de plaques ateroscleròtiques de la metal·loproteïnasa de matriu 9 [183] i del factor de creixement de l'endoteli vascular [184]. Els resultats poden ser rellevants ja que indiquen quins són els biomarcadors més prometedors, als quals idealment haurien de ser destinats els recursos de la futura recerca, sempre que el benefici potencial de la nova evidència superi els costos de la investigació. A més a més, donat que els pacients amb MRC i altres comorbiditats —com diabetis o dislipèmies— són el

subgrup amb major risc, aquests podrien beneficiar-se d'un cribratge de nivells alts d'aquests biomarcadors per a posteriors intervencions preventives, o inclús, també, d'un futurible tractament preventiu centrat en la disminució dels nivells d'aquests biomarcadors. Aquestes hipotètiques intervencions serien més eficients en aquest subgrup de pacients que en altres, pel fet de requerir un menor nombre de pacients a tractar per evitar un esdeveniment cardiovascular, la qual cosa suposaria un major benefici net en salut per una determinada disposició a pagar.

En el quart treball d'aquesta tesi s'ha implementat un model analític semi-Markov per avaluar l'impacte en costos i en salut de l'augment en l'ús de la ICP en pacients ancians amb SCA. El model conceptualitza el curs de la malaltia en un horitzó temporal de 8 anys des de l'ingrés hospitalari. Els resultats de l'estudi indiquen que augmentar el nombre de pacients ancians que es beneficiïn de la ICP durant l'ingrés per SCA és rendible en un llinyar de disposició a pagar molt inferior als estàndards en països europeus. Si bé tots els escenaris van proporcionar resultats d'eficiència molt similars, l'escenari de proporcionar la ICP a tots els pacients amb la menor probabilitat de rebre-la en la pràctica clínica habitual va ser el més rendible en homes. En les dones, els resultats cost-utilitat van ser gairebé els mateixos en els escenaris de proporcionar la ICP a pacients amb la menor i la major probabilitat de beneficiar-se'n en l'actual pràctica clínica. El fet que un dels escenaris més rendibles fos proporcionar ICP a tots els pacients amb la probabilitat més baixa de beneficiar-se de la intervenció pot indicar que la decisió de proporcionar-la depèn de molts factors diferents, alguns dels quals probablement no van ser inclosos en les estimacions de les reduccions de mortalitat per ICP, i que no sempre es poden avaluar.

El model de *propensity score* utilitzat per calcular la probabilitat de rebre la ICP es va basar en l'edat, el tipus d'esdeveniment d'SCA, l'aparició d'infart de miocardi previ i la presència de diabetis, hipertensió i símptomes d'insuficiència cardíaca. El model va mostrar una capacitat de discriminació del 70%, la qual cosa indica que no es pot definir totalment la indicació de la ICP amb les covariables disponibles. Tenint en compte això, podríem esperar que una fracció de pacients classificats amb una baixa probabilitat de rebre la ICP podria tenir en realitat una probabilitat més alta de rebre-la. Les directrius clíniques actuals recomanen avaluar les comorbiditats, l'esperança de vida, la qualitat de vida, la fragilitat i les preferències del pacient, per tal de decidir la indicació d'una estratègia invasiva en pacients grans [146, 147]. Tot i això, aquests paràmetres no sempre s'utilitzen per prendre decisions terapèutiques, i diversos estudis, així com el nostre, suggereixen que alguns pacients amb SCA d'alt risc reben probablement un tractament més conservador del que haurien de rebre [185-187].

4.3 Limitacions

Donat que els quatre treballs de la tesi tenen dissenys i fonts de dades diferents, les principals limitacions són les pròpies de cadascun dels treballs. Els

dos treballs d'avaluació econòmica basats en models analítics comparteixen la limitació pròpia de les fonts d'evidència utilitzades i dels supòsits metodològics. En ambdós models, s'ha fet servir la major evidència disponible per a la població objecte d'estudi, adaptant els supòsits metodològics necessaris per tal de poder-la incorporar al model. De tota manera, se'n va avaluar detingudament la incertesa. En l'avaluació de les estratègies de cribratge de càncer de mama, la incertesa va ser avaluada mitjançant anàlisi de sensibilitat determinista univariable dels paràmetres de costos i d'AVAQ, per ser els paràmetres amb major impacte en els resultats, tal com ja es va mostrar en treballs previs [188, 189], i mitjançant l'anàlisi dels nous supòsits incorporats en la modelització del treball inclòs en aquesta tesi: la distribució dels grups de risc considerats al model i el percentatge de sobrediagnòstic de càncer de mama invasiu i *in situ*. L'anàlisi de sensibilitat probabilística va ser desestimada, donat que, en tractar-se d'un model semi-Markov basat en equacions diferencials que cal solucionar numèricament, el cost computacional era inassumible. Tanmateix, es va validar el calibratge del model comparant els resultats obtinguts amb una revisió sistemàtica de Cochrane [190], la revisió independent de l'Independent UK Panel on Breast Cancer Screening [191], la revisió integral del programes de cribratge a Europa del grup de treball EUROSCREEN [192] i l'estudi INCA (INterval CAncer) [193] o altres estudis de càncer d'interval [194, 195].

En l'avaluació de l'ús de la ICP en pacients ancians, la incertesa estructural va ser avaluada mitjançant la implementació de diversos escenaris d'increment de l'ús de la ICP per tercils de probabilitat (*propensity score*) de rebre la intervenció en la pràctica clínica habitual. La incertesa paramètrica va ser avaluada mitjançant anàlisi de sensibilitat determinista univariable i anàlisi de sensibilitat probabilística. El model va mostrar ser molt robust i va proporcionar estimacions molt precises de l'eficiència dels diferents escenaris, probablement com a conseqüència de la potència dels estudis clínics, dels quals es van extreure les estimacions dels paràmetres i els errors estàndards. No obstant això, el pronòstic dels pacients després de l'alta hospitalària no es va incorporar a l'anàlisi de sensibilitat, donat que l'estimació de les probabilitats de recurrència i de mort per als cicles del model només pot ser obtinguda per mètodes d'integració numèrica basats en *bootstrap* computacionalment intensius, cosa que en fa inviable la inclusió en l'anàlisi de sensibilitat. Hi ha altres limitacions en aquest estudi que convé tenir en compte. En primer lloc, no es va avaluar el cost-utilitat segons si els pacients amb SCA presentaven o no elevació del segment ST. Les estimacions del pronòstic extrahospitalari utilitzades en aquest estudi es van obtenir d'una base de dades sanitària validada del nord-est d'Espanya que no tenia informació sobre l'elevació del segment ST. Tanmateix, tenint en compte que tots els pacients amb elevació del segment ST i la majoria de pacients amb SCA sense elevació del segment ST són candidats a rebre una ICP, l'avaluació realitzada en aquesta tesi proporciona una primera aproximació al cost-utilitat d'augmentar-ne l'ús en pacients ancians. En segon lloc, és possible que la població de l'estudi no reflecteixi l'actual pràctica clínica dels pacients amb SCA als països europeus inclosos en l'avaluació econòmica realitzada. Tot i que les estimacions de la

probabilitat de rebre la ICP i de la mortalitat hospitalària es van obtenir d'una base de dades gran de pacients amb SCA a Europa (EUROTRACS) [156], no tots els pacients corresponen a ingressos durant la darrera dècada. Malgrat que esperaríem un augment de l'ús de la ICP i una reducció de la mortalitat hospitalària en dades més contemporànies, estudis amb dades més recents mostren que l'ús de la ICP no ha augmentat en pacients ancians al mateix ritme que en pacients més joves, tot i que els motius romanen incerts [196]. En tercer lloc, les probabilitats de recurrència i mortalitat després de l'alta hospitalària es van estimar a partir d'una font validada a Espanya i es van assumir per als altres 5 països. Tot i que se sap que hi ha diferències en la gestió clínica entre els països europeus [197], es podria esperar que la història natural de la malaltia fos similar entre països.

Una limitació en l'anàlisi conjunta de les mesures longitudinals de densitat mamària i risc de càncer de mama és la limitada grandària de la cohort d'estudi per dur a terme una anàlisi de factors de risc de càncer. Però el nostre model no pretén ser una eina d'ús habitual, sinó l'exemple d'aplicació pràctica d'una metodologia que podria emprar-se per obtenir mesures més precises del risc individual de càncer de mama. Per això, caldrien estudis de cohorts més grans de base poblacional, amb disposició de mesures longitudinals de la densitat mamària, així com dels altres factors de risc coneguts. Tot i que el nostre model incorpora els antecedents familiars i els procediments previs de mama, com per exemple els models de Chen *et al.* [167] i de Barlow *et al.* [168], els models actuals incorporen les lesions benignes [171] i també el risc poligènic [174], no disponibles en les dades del nostre estudi. Les assumpcions metodològiques i les alternatives de la modelització estan extensament discutides en la publicació del segon treball d'aquesta tesi.

En el tercer treball d'aquesta tesi, el petit nombre d'esdeveniments cardiovasculars pot haver limitat la capacitat d'identificar més clarament els predictors de risc cardiovascular. A més a més, cal tenir present que, tot i que la recerca de biomarcadors sèrics amb capacitat predictiva potencial d'esdeveniments cardiovasculars és un camp d'estudi molt actiu, un nombre molt reduït arriba a la clínica [198, 199]. Per superar aquestes limitacions s'ha fet servir un model de bosc aleatori de supervivència en presència d'esdeveniments competius, que ha proporcionat una ordenació de la capacitat predictiva de tots els predictors de manera simultània. En general, els algorismes d'aprenentatge automàtic permeten identificar predictors forts en situacions amb una relació senyal-soroll baixa, característica que els fa especialment útils en la cerca de nous biomarcadors.

4.4 Fortaleses

La principal fortalesa, potser comuna a tots els treballs que configuren aquesta tesi, és l'aplicació adequada i rigorosa dels mètodes matemàtics i estadístics necessaris per proporcionar evidència útil per a la presa de decisions, segons els objectius plantejats en els diferents treballs. En el primer treball es va utilitzar un model molt detallat que va permetre avaluar l'eficiència i

les relacions dany-benefici de 2.625 escenaris de cribratge diferents, ja sigui basats en el risc o no. Només amb un model analític és possible fer una avaluació tan exhaustiva. Amb un elevat nombre de citacions (69 a Scopus a 9 de juny de 2020), i més enllà del reconeixement de la qualitat del nostre model en una revisió sistemàtica realitzada per un investigador independent [165], el nostre treball s'incorpora a tota l'evidència publicada sobre els beneficis del cribratge adaptat al risc, majoritàriament avaluada en models analítics. Potser també degut a això, en els darrers anys s'han endegat assajos clínics per avaluar l'eficàcia, la seguretat i l'acceptabilitat del cribratge basat en el risc en comparació amb l'estratègia uniforme estàndard, com és el cas de l'estudi WISDOM als Estats Units [200, 201], el multicèntric MyPeBS (*My Personalized Breast Screening*) a Europa [202] i el TBST (*Tailored Breast Screening Trial*) a Itàlia [203]. S'espera que els resultats d'aquests estudis corroborin les conclusions dels de modelització i consolidin l'evidència sobre la major efectivitat del cribratge personalitzat del càncer de mama.

A diferència de treballs anteriors, el nostre estudi de modelització conjunta ha estat el primer que ha utilitzat la història longitudinal completa de la densitat mamària per avaluar el risc de càncer de mama al llarg del temps, tant poblacional com individual. La flexibilitat de la modelització emprada permet obtenir prediccions individuals del risc de càncer de mama, donada tota l'evolució de la densitat mamària, així com prediccions longitudinals individuals de valors futurs de densitat mamària. De fet, la nostra proposta podria aplicar-se a una varietat de problemes reals per analitzar l'impacte de l'evolució de variables endògenes ordinals en el risc d'experimentar l'esdeveniment d'interès.

Potser una de les fortaleeses més grans de l'estudi dels biomarcadors cardiovasculars és el fet d'haver considerat un nombre elevat de biomarcadors potencials alhora, i juntament amb biomarcadors i factors de risc coneguts. La contribució a la capacitat de discriminació va poder ser avaluada de manera conjunta gràcies a l'algorisme de bosc aleatori, que a part de ser més fàcilment interpretable que altres algorismes d'aprenentatge automàtic alternatiu, com ara les màquines de vector de suport (*support-vector machines*), les xarxes neuronals artificials (*artificial neural networks*) o certs algorismes *boosting*, acostuma a proporcionar resultats amb altes precisions i sovint requereix menys temps computacional.

El model emprat en l'avaluació de la ICP en pacients ancians va mostrar una robustesa estructural molt elevada, reflectida en l'alta coherència dels resultats per als diferents escenaris considerats. Malgrat els amplis intervals tinguts en compte per a paràmetres amb dispersió desconeguda, el nostre model també va proporcionar resultats molt fiables, probablement a causa de l'ús de grans cohorts de pacients amb ACS per obtenir els paràmetres del model. L'estimació de l'efectivitat, en termes de reducció en la mortalitat hospitalària associada a la utilització de la ICP, es va obtenir a partir de les dades de tots els pacients ancians ingressats per SCA durant els anys 2000-2014 als hospitals europeus participants en el projecte EUROTRACS [156]. Les estimacions del risc de recurrència i mortalitat després de l'alta hospitalària

es van obtenir a partir de dades poblacionals del Sistema d'Informació per al Desenvolupament de la Recerca en Atenció Primària (SIDIAP) [204, 205], seleccionant els pacients amb esdeveniment d'SCA primària durant el període 2007-2013. Per tant, considerem que la representativitat de la població analitzada reforça l'alt nivell d'evidència aportada pels resultats obtinguts del model analític, que indiquen l'eficiència d'estendre l'ús de la ICP a tots els pacients ancians ingressats per SCA. A més a més, com que el projecte EUROTRACS també va recopilar dades de costos sanitaris directes dels sis països participants, es van poder reportar resultats adaptats als diferents sistemes sanitaris, la qual cosa facilita la presa de decisions segons la perspectiva de les diferents jurisdiccions.

4.5 Futures línies de recerca

Aquesta tesi s'emmarca dins la línia d'avaluació econòmica i salut de la Universitat de Lleida. En l'era de la medicina personalitzada, la línia de recerca té com a principal objectiu la personalització del cribratge de càncer de mama, impulsada per diversos factors: l'evidència disponible sobre els efectes adversos, en especial el sobrediagnòstic; la disseminació de la medicina personalitzada; la presa de decisions compartides metge-pacient; el desenvolupament de models de risc de càncer de mama més precisos, i la possibilitat de realitzar intervencions preventives en dones d'alt risc. Sense cap tipus de dubte, l'avaluació econòmica del cribratge presentada en aquesta tesi proporciona un dels ingredients fonamentals, sobre el qual s'ha continuat avançant en els darrers anys amb l'estudi InforMa [206]. L'estudi InforMa és un RCT dut a terme pel nostre grup de recerca que ha permès evidenciar que el fet d'aportar informació quantitativa sobre beneficis i danys del cribratge ha produït un augment considerable del coneixement i la presa de decisions informada, amb una alta acceptació dels materials informatius [207]. Actualment, està en marxa l'estudi DECIDO, que pretén avaluar l'acceptabilitat i la viabilitat d'oferir cribratge personalitzat del càncer de mama dins el sistema públic de salut [208]. Entre els seus objectius, s'espera que l'estudi DECIDO proporcioni informació sobre les barreres i els facilitadors del canvi de paradigma del cribratge uniforme al cribratge personalitzat, fet que implica diversos nivells del sistema sanitari que haurien de coordinar-se de la manera més eficient possible, així com l'acceptabilitat entre les dones.

El desenvolupament metodològic que permeti donar resposta als problemes de salut plantejats és també focus d'interès principal del nostre grup de recerca. El model conjunt proposat en el desenvolupament d'aquesta tesi és fruit de les dues vessants principals de la nostra línia de recerca. Ara bé, per una banda, per tal de proporcionar una eina útil, a la pràctica, de mesura del risc individual de càncer de mama, el model proposat hauria de ser ajustat a una població més gran, potser també més representativa, amb un seguiment prou llarg i recopilant les dades de tots els factors de risc rellevants. Malauradament, el disseny d'una nova cohort prospectiva amb aquest objectiu seria inabastable per a nosaltres. Si els programes de

cribratge poblacional registressin la densitat mamària de manera rutinària en cada mamografia es podrien fer estudis amb milers de dones i múltiples mesures longitudinals. A més a més, una anàlisi d'una simple mostra de saliva permetria identificar les variants gèniques conegudes que configuren el factor de risc poligènic incorporat als models de risc més recents. Per una altra banda, el càlcul de la mesura de discriminació del model conjunt seguint un enfocament purament bayesià és un aspecte d'alt interès metodològic per a nosaltres i per a la comunitat científica. A més a més, és necessari mesurar la millora en la capacitat predictiva respecte d'altres modelitzacions que no inclouen les trajectòries longitudinals dels factors de risc per tal de justificar el major esforç que suposa la modelització conjunta de mesures longitudinals i temps fins a l'esdeveniment. Rizopoulos va proposar una metodologia pseudobayesiana per al càlcul de la capacitat predictiva d'un model conjunt [107], que seria un bon punt de partida per obtenir una primera estimació per al nostre model.

En general, qualsevol problema d'avaluació econòmica i/o estudis amb dissenys que necessiten metodologia estadística avançada representen oportunitats de col·laboració amb altres grups de recerca en ciències de la salut per avançar en el coneixement, sempre amb l'objectiu de millorar la salut i la qualitat de vida de les persones. Els altres dos treballs inclosos en aquesta tesi en són un bon exemple. L'apèndix A incorpora una relació d'altres col·laboracions amb objectius d'avaluació econòmica, o de modelització de resultats en salut.

Capítol 5

Conclusions

Al llarg dels treballs inclosos en aquesta tesi ha quedat palès el paper i la utilitat de la modelització matemàtica i estadística per guiar les decisions, la planificació i la gestió en salut. A continuació s'enumeren les principals conclusions de la tesi, segons el problema de decisió plantejat.

Quant a l'avaluació dels programes de cribratge de càncer de mama:

1. Les estratègies de cribratge de càncer de mama basades en el risc proporcionen major benefici net en salut que les estratègies uniformes estàndards.
2. Cal desenvolupar mesures precises de risc individual de càncer de mama i avançar en l'organització de programes de cribratge basats en el risc.

En relació amb la incorporació de mesures longitudinals de densitat mamària per a l'estimació de risc de càncer de mama mitjançant models conjunts:

3. Afegir la trajectòria temporal de la densitat mamària als actuals models de risc pot proporcionar estimacions individuals del risc de càncer de mama més precises i permetre avançar cap a la personalització del cribratge.
4. Es podria utilitzar un model conjunt similar al proposat, que ha utilitzat la història longitudinal completa de la densitat mamària, per a la vigilància del risc de càncer de mama, per programar exàmens de cribratge a partir de prediccions dinàmiques individuals i també per discutir estratègies de prevenció per a dones amb risc elevat.

Pel que fa a l'avaluació de la capacitat de discriminació de nous biomarcadors de risc cardiovascular en pacients amb MRC:

5. La recerca futura hauria de focalitzar-se en els biomarcadors amb major capacitat predictiva —en aquest treball, bàsicament l'osteopontina—, per davant d'aquells que mostren una contribució quasi nul·la en la predicció cardiovascular en pacients amb MRC.

6. Els pacients amb MRC i altres comorbiditats —diabetis, dislipèmia, hipertensió— podrien beneficiar-se d'un cribratge de nivells alts d'aquests biomarcadors.

Quant a l'avaluació econòmica de l'increment de l'ús de la ICP en població anciana amb SCA:

7. En comparació amb la pràctica clínica habitual, realitzar la ICP a una proporció més alta de pacients ancians ingressats per SCA seria eficient, independentment de l'estratègia seleccionada. Per tant, el cost-utilitat no hauria de ser un motiu per no incrementar l'ús de la ICP en aquesta població.
8. El fet que sigui tant eficient proporcionar ICP a pacients ancians amb SCA amb baixa probabilitat de beneficiar-se de la intervenció en la pràctica clínica actual com proporcionar-la a pacients amb alta probabilitat de beneficiar-se'n en la pràctica clínica actual, suggereix que els criteris d'aplicació de la ICP en pacients ancians haurien de revisar-se i definir-se amb més claredat.

5.1 Implicacions per a la salut pública

Els treballs d'aquesta tesi han incorporat evidència en les seves respectives àrees, fet que facilita la presa de decisions en diferents nivells del sistema de salut i en el camp de la recerca.

Amb els diversos i excel·lents treballs publicats sobre el benefici de la personalització del cribratge de càncer de mama, entre els quals es troba el nostre, i en l'era de la medicina personalitzada, s'espera que el sistema públic de salut emprengui les accions oportunes per transitar cap a estratègies de cribratge personalitzades, i que faciliti informació a les dones sobre els beneficis i els efectes adversos per a una presa de decisions adequada sobre la seva participació. L'obligació dels gestors i decisors públics és informar la població i destinar els recursos de la manera més eficient possible. El cribratge personalitzat és l'alternativa eficient, alhora que redueix els costos de les estratègies uniformes actuals.

L'avaluació econòmica de la ICP s'afegeix als resultats d'efectivitat reportats per l'estudi EUROTRACS amb ànim d'encoratjar gestors i clínics a estendre l'ús de la ICP a pacients ancians, que podrien beneficiar-se d'una intervenció que sovint no se'ls aplica. En vista dels resultats del nostre treball, la disseminació de la ICP proporcionaria un benefici net en salut enorme acceptant una disposició a pagar estàndard en països de l'Europa occidental.

En l'era de les dades massives (*big data*), també la salut pública ha de beneficiar-se d'un paradigma caracteritzat per un gran volum de dades, d'alta complexitat i generades a gran velocitat. Això fa necessaris mètodes d'estimació eficients utilitzant conjunts de dades d'ús de recursos sanitaris, dades clíniques, dades de dispositius personals i moltes altres fonts. Els models de regressió tradicionals sovint no són adequats. Els algorismes d'aprenentatge

automàtic aborden i superen aquestes limitacions de manera eficaç, tot i que estan subjectes als potencials biaixos habituals dels estudis observacionals, i per tant han de ser avaluats amb el mateix rigor que els models estadístics convencionals. Tanmateix, la intel·ligència artificial i l'aprenentatge automàtic també tenen limitacions que cal tenir en compte abans d'adoptar àmpliament la tecnologia en la presa de decisions en salut. Tal com s'apunta en un article publicat al *New England Journal of Medicine* el desembre de 2019, «com totes les intervencions mèdiques, implementar l'aprenentatge automàtic per al diagnòstic de càncer tindrà alguns beneficis i alguns perjudicis. Pot augmentar la velocitat i la consistència del diagnòstic, però també pot agreujar el sobrediagnòstic» [209]. Per tot plegat, és del tot necessari que els analistes acreditin una sòlida formació en ciències i tècniques estadístiques.

Finalment, els recursos destinats a la investigació tenen el seu impacte en la salut de les persones a través de l'elecció dels projectes de recerca finançats, ja sigui per fons públics o privats. Les innovacions han de ser de prou valor afegit per al finançador per compensar el cost de la recerca. Així, per exemple, els estudis de nous biomarcadors són útils per centrar els esforços futurs en aquells que hagin mostrat major associació amb el resultat d'interès. La innovació metodològica en l'anàlisi de dades també contribueix en l'avenç del coneixement, ja que omple els buits existents en les tècniques estadístiques, que es van creant a mesura que altres disciplines avancen o per l'aparició de noves tecnologies. Per tant, destinar recursos per al desenvolupament de metodologies estadístiques que poden aportar solucions a problemes oberts també té un potencial impacte positiu en la salut de les persones. Aquest seria el cas, per exemple, de la modelització conjunta de dades longitudinals i temps fins a l'esdeveniment aplicada al desenvolupament de mesures dinàmiques del risc individual de càncer de mama. Un model amb la història longitudinal de la densitat mamària que inclogués els altres factors de risc coneguts, com les lesions benignes i el risc poligènic, podria assolir capacitats predictives aplicables a la personalització del cribratge de càncer de mama, o a l'establiment d'estratègies de prevenció per a persones amb risc elevat.

Apèndix A

Altres articles amb resultats col·laterals

Durant el desenvolupament d'aquesta tesi també s'han publicat, amb la participació del doctorand, altres treballs d'avaluació econòmica o de modelitzacions de resultats en salut per aplicar en posteriors avaluacions econòmiques o en la pràctica clínica. La taula A.1 mostra les referències d'aquests treballs amb llurs objectius.

TAULA A.1: Objectius col·laterals a la tesi i referències a les publicacions originals

Objectiu	Article
Avaluar el calibratge i la capacitat de discriminació de tres models predictius de risc de càncer de mama (Gail, Chen, Barlow)	Arrospide <i>et al.</i> BMC Cancer 13:587
Avaluar l'impacte dels biaixos en la supervivència del càncer de mama deguts a la detecció precoç (<i>lead-time bias</i> i <i>length-sampling bias</i>)	Roso-Llorach <i>et al.</i> SORT 38:139-160
Determinar utilitats en salut genèriques per a pacients espanyols amb malaltia pulmonar obstructiva crònica estratificades segons diferents classificacions (GOLD 2007, GOLD 2013, GesEPOC 2012, índex BODEx)	Miravittles <i>et al.</i> Health and Quality of Life Outcomes 12:120

Continua a la pàgina següent...

...Continua de la pàgina anterior.

Objectiu	Article
Determinar les variables demogràfiques i clíniques significativament associades a les utilitats en salut per a pacients amb malaltia pulmonar obstructiva crònica	Miravittles <i>et al.</i> International Journal of Chronic Obstructive Pulmonary Disease 10:367-377
Determinar la rendibilitat d'omalizumab en comparació amb el tractament estàndard per al control de l'asma persistent greu segons dades de pacients tractats en una unitat d'asma especialitzada	Vennera <i>et al.</i> Clinical Drug Investigation 36:567-578
Anàlisi de minimització de costos de l'onabotulinumtoxina A i l'abobotulinumtoxina A, partint de dosis reals administrades a nens amb espasticitat associada a la deformitat dinàmica del peu equí deguda a paràlisi cerebral	Tapias <i>et al.</i> Farmàcia Hospitalària 40:412-426
Estimar els costos sanitaris directes del tractament antimicrobià domiciliari endovenós administrat per les unitats d'hospitalització domiciliària a Espanya	González-Ramallo <i>et al.</i> International Journal of Antimicrobial Agents 50:114-118
Avaluar si la malaltia renal crònica promou un perfil de lípids proaterogènic modificant la composició de les lipoproteïnes i el nombre de partícules	Bermudez-López <i>et al.</i> Expert Opinion on Therapeutic Targets 23:619-630

A.1 Article A1

Títol: An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain

Autors: Arrospide A, Forné C, Rué M, Torà N, Mar J, Baré M

Revista: BMC Cancer. 2013 Dec 10;13:587

Factor d'impacte de l'any de la publicació: 3,319 (percentil 63,8 a *Oncology*)

DOI: 10.1186/1471-2407-13-587

Resum original:

Background: The aim of this study was to evaluate the calibration and discriminatory power of three predictive models of breast cancer risk.

Methods: We included 13,760 women who were first-time participants in the Sabadell-Cerdanyola Breast Cancer Screening Program, in Catalonia, Spain. Projections of risk were obtained at three and five years for invasive cancer using the Gail, Chen and Barlow models. Incidence and mortality data were obtained from the Catalan registries. The calibration and discrimination of the models were assessed using the Hosmer-Lemeshow C statistic, the area under the receiver operating characteristic curve (AUC) and the Harrell's C statistic.

Results: The Gail and Chen models showed good calibration while the Barlow model overestimated the number of cases: the ratio between estimated and observed values at 5 years ranged from 0.86 to 1.55 for the first two models and from 1.82 to 3.44 for the Barlow model. The 5-year projection for the Chen and Barlow models had the highest discrimination, with an AUC around 0.58. The Harrell's C statistic showed very similar values in the 5-year projection for each of the models. Although they passed the calibration test, the Gail and Chen models overestimated the number of cases in some breast density categories.

Conclusions: These models cannot be used as a measure of individual risk in early detection programs to customize screening strategies. The inclusion of longitudinal measures of breast density or other risk factors in joint models of survival and longitudinal data may be a step towards personalized early detection of BC.

Keywords: Breast cancer, Screening, Risk models, Individual risk, Breast density

A.2 Article A2

Títol: Assessing the impact of early detection biases on breast cancer survival of Catalan women

Autors: Roso-Llorach A, Forné C, Macià F, Galceran J, Marcos-Gragera R, Rué M

Revista: SORT - Stat Oper Res T. 2014 Jul-Dec;38:139-160

Factor d'impacte de l'any de la publicació: 1,333 (percentil 69,3 a *Statistics & Probability*)

DOI: 10.2436/20.8080.02.8

Resum original:

Survival estimates for women with screen-detected breast cancer are affected by biases specific to early detection. Lead-time bias occurs due to the advance of diagnosis, and length-sampling bias because tumors detected on screening exams are more likely to have slower growth than tumors symptomatically detected. Methods proposed in the literature and simulation were used to assess the impact of these biases. If lead-time and length-sampling biases were not taken into account, the median survival time of screen-detected breast cancer cases may be overestimated by 5 years and the 5-year cumulative survival probability by between 2.5 to 5 percent units.

Keywords: Breast cancer, early detection, screening, lead time bias, length bias, survival.

A.3 Article A3

Títol: Generic utilities in chronic obstructive pulmonary disease patients stratified according to different staging systems

Autors: Miravittles M, Huerta A, Fernández-Villar JA, Alcázar B, Villa G, Forné C, Cuesta M, Crespo C, García-Río F

Revista: Health Qual Life Outcomes. 2014 Sep 5;12:120

Factor d'impacte de l'any de la publicació: 2,120 (percentil 63,1 a *Health Care Sciences & Services*)

DOI: 10.1186/s12955-014-0120-5

Resum original:

Background: To determine generic utilities for Spanish chronic obstructive pulmonary disease (COPD) patients stratified by different classifications: GOLD 2007, GOLD 2013, GesEPOC 2012 and BODEx index.

Methods: Multicentre, observational, cross-sectional study. Patients were aged ≥ 40 years, with spirometrically confirmed COPD. Utility values were derived from EQ-5D-3 L. Means, standard deviations (SD), medians and interquartile ranges (IQR) were computed based on the different classifications. Differences in median utilities between groups were assessed by non-parametric tests.

Results: 346 patients were included, of which 85.5% were male with a mean age of 67.9 (SD=9.7) years and a mean duration of COPD of 7.6 (SD=5.8) years; 80.3% were ex-smokers and the mean smoking history was 54.2 (SD=33.2) pack-years. Median utilities (IQR) by GOLD 2007 were 0.87 (0.22) for moderate; 0.80 (0.26) for severe and 0.67 (0.42) for very-severe patients ($p < 0.001$ for all comparisons). Median utilities by GOLD 2013 were group A: 1.0 (0.09); group B: 0.87 (0.13); group C: 1.0 (0.16); group D: 0.74 (0.29); comparisons were statistically significant ($p < 0.001$) except A vs C. Median utilities by GesEPOC phenotypes were 0.84 (0.33) for non exacerbator; 0.80 (0.26) for COPD-asthma overlap; 0.71 (0.62) for exacerbator with emphysema; 0.72 (0.57) for exacerbator with chronic bronchitis ($p < 0.001$). Comparisons between patients with or without exacerbations and between patients with COPD-asthma overlap and exacerbator with chronic bronchitis were statistically-significant ($p < 0.001$). Median utilities by BODEx index were: group 0–2: 0.89 (0.20); group 3–4: 0.80 (0.27); group 5–6: 0.67 (0.29); group 7–9: 0.41 (0.31). All comparisons were significant ($p < 0.001$) except between groups 3–4 and 5–6.

Conclusion: Irrespective of the classification used utilities were associated to disease severity. Some clinical phenotypes were associated with worse utilities, probably related to a higher frequency of exacerbations. GOLD

2007 guidelines and BODEx index better discriminated patients with a worse health status than GOLD 2013 guidelines, while GOLD 2013 guidelines were better able to identify a smaller group of patients with the best health.

Keywords: Chronic obstructive pulmonary disease, Health utility, Health related quality of life, GOLD, GesEPOC

A.4 Article A4

Títol: Clinical variables impacting on the estimation of utilities in chronic obstructive pulmonary disease

Autors: Miravittles M, Huerta A, Valle M, García-Sidro P, Forné C, Crespo C, López-Campos JL

Revista: Int J Chron Obstruct Pulmon Dis. 2015 Feb 16;10:367-377

Factor d'impacte de l'any de la publicació: 3,046 (percentil 71,5 a *Respiratory System*)

DOI: 10.2147/COPD.S76397

Resum original:

Purpose: Health utilities are widely used in health economics as a measurement of an individual's preference and show the value placed on different health states over a specific period. Thus, health utilities are used as a measure of the benefits of health interventions in terms of quality-adjusted life years. This study aimed to determine the demographic and clinical variables significantly associated with health utilities for chronic obstructive pulmonary disease (COPD) patients.

Patients and methods: This was a multicenter, observational, cross-sectional study conducted between October 2012 and April 2013. Patients were aged ≥ 40 years, with spirometrically confirmed COPD. Utility values were derived from the preference-based generic questionnaire EQ-5D-3L applying weighted Spanish societal preferences. Demographic and clinical variables associated with utilities were assessed by univariate and multivariate linear regression models.

Results: Three hundred and forty-six patients were included, of whom 85.5% were male. The mean age was 67.9 (standard deviation [SD]=9.7) years and the mean forced expiratory volume in 1 second (%) was 46.2% (SD=15.5%); 80.3% were former smokers, and the mean smoking history was 54.2 (SD=33.2) pack-years. Median utilities (interquartile range) were 0.81 (0.26) with a mean value of 0.73 (SD=0.29); 22% of patients had a utility value of 1 (ceiling effect) and 3.2% had a utility value lower than 0. The factors associated with utilities in the multivariate analysis were sex (beta= -0.084, 95% confidence interval [CI]: -0.154; -0.013 for females), number of exacerbations the previous year (-0.027, 95% CI: -0.044; -0.010), and modified Medical Research Council Dyspnea Scale (mMRC) score (-0.123 [95% CI: -0.185; -0.061], -0.231 [95% CI: -0.301; -0.161], and -0.559 [95% CI: -0.660; -0.458] for mMRC scores 2, 3, and 4 versus 1), all $p < 0.05$.

Conclusion: Multivariate analysis showed that female sex, frequent exacerbations, and an increased level of dyspnea were the main factors associated with reduced utility values in patients with COPD.

Keywords: COPD, health utility, health-related quality of life, multivariate linear regression

A.5 Article A5

Títol: Cost-Effectiveness Analysis of Omalizumab for the Treatment of Severe Persistent Asthma in Real Clinical Practice in Spain

Autors: Vennera MC, Valero A, Uría E, Forné C, Picado C

Revista: Clin Drug Investig. 2016 Jul;36(7):567-578

Factor d'impacte de l'any de la publicació: 1,853 (percentil 31,7 a *Pharmacology & Pharmacy*)

DOI: 10.1007/s40261-016-0402-2

Resum original:

Background and Objective: Omalizumab is a humanized monoclonal antibody that targets circulating immunoglobulin E molecules to treat severe uncontrolled asthma. The aim of this study was to determine the cost effectiveness of omalizumab compared with standard treatment for the control of severe persistent asthma according to data from patients treated in a specialized asthma unit.

Methods This was an observational, retrospective, single-center study in the setting of the Pulmonology and Respiratory Allergy Service, Thorax Institute, Hospital Clínic de Barcelona, Barcelona, Spain. Data were collected by review of medical records of 86 uncontrolled severe persistent asthma patients treated with omalizumab from January 2005 to April 2014. Effectiveness was assessed by the reduction in asthma exacerbations and 3-point increases in the Asthma Control Test (ACT) score. The economic evaluation was performed from the societal perspective, including direct health costs (resource use and drug treatments) and indirect costs (disease impact on labor productivity) in 2016 Euros. The time horizon was 12 months before and after the initiation of treatment with omalizumab. Results were expressed using the incremental cost-effectiveness ratio (ICER).

Results: Taking into account only direct costs, the ICERs were €1487.46 (95% confidence interval [CI] 1241.21-1778.34) per exacerbation avoided and €5425.13 (95% CI 4539.30-6551.03) per 3-point increase in the ACT. When indirect costs were included, the ICERs were €1130.93 (95% CI 909.08-1392.86) per exacerbation avoided, and €4124.79 (95% CI 3281.69-5186.73) per 3-point increase in the ACT.

Conclusions: The results of this study confirm the effectiveness of the addition of omalizumab to standard therapy in patients with uncontrolled severe persistent asthma.

Keywords: Asthma, Indirect Cost, Salmeterol, Asthma Control, Asthma Exacerbation

A.6 Article A6

Títol: Cost-minimization analysis in the treatment of spasticity in children with cerebral palsy with botulinum toxin type A: an observational, longitudinal, retrospective study

Autors: Tapias G, García-Romero M, Crespo C, Cuesta M, Forné C, Pascual-Pascual SI

Revista: Farm Hosp. 2016 Sep 1;40(5):412-426

SCImago Journal Rank (SJR) de l'any de la publicació: 0,227 (tercer quartil a *Medicine (miscellaneous)*)

DOI: 10.7399/fh.2016.40.5.10429

Resum original:

Objective: Cost-minimization analysis of *onabotulinumtoxinA* and *abobotulinumtoxinA*, taking into account the real dose administered to children with spasticity associated with dynamic equinus foot deformity due to cerebral palsy.

Method: A single centre, observational, longitudinal, and retrospective study which included spastic paediatric patients aged 2-to-18-years and treated with *onabotulinumtoxinA* or *abobotulinumtoxinA* from December 1995 to October 2012, in the Paediatric Neurology Unit of a first-level Spanish hospital. A longitudinal analysis of spasticity severity was made to confirm the similar effectiveness of both treatments. Cost minimization was analyzed using the dose administered and the direct costs (pharmacological and medical visits costs) from the perspective of the National Health System (in 2016 euros).

Results: We analyzed 895 patients with paediatric spasticity: 543 were treated only with *onabotulinumtoxinA*, 292 only with *abobotulinumtoxinA*, and 60 with both treatments. The mean doses administered were 5.44 U/kg (SD=2.17) for *onabotulinumtoxinA*, and 14.73 U/kg (5.26) for *abobotulinumtoxinA*. The total annual direct cost (pharmacological and medical visits) was €839.56 for *onabotulinumtoxinA* and €631.23 for *abobotulinumtoxinA*, which represents a difference of €208.34 per year in favour of treatment with *abobotulinumtoxinA*.

Conclusions: It has been shown that in real clinical practice, the cost per patient and year for treatment of paediatric spasticity was lower when *abobotulinumtoxinA* was used.

Keywords: Botulinum toxins; Paediatric spasticity; Clinical practice dose; Equinus foot; Spain

A.7 Article A7

Títol: Costs of outpatient parenteral antimicrobial therapy (OPAT) administered by Hospital at Home units in Spain

Autors: González-Ramallo VJ, Mirón-Rubio M, Mujal A, Estrada O, Forné C, Aragón B, Rivera AJ

Revista: Int J Antimicrob Agents. 2017 Jul;50(1):114-118

Factor d'impacte de l'any de la publicació: 4,253 (percentil 87,2 a *Pharmacology & Pharmacy*)

DOI: 10.1016/j.ijantimicag.2017.02.017

Resum original:

The aim of this study was to assess the direct healthcare costs of outpatient parenteral antimicrobial therapy (OPAT) administered by Hospital at Home (HaH) units in Spain. An observational, multicentre, economic evaluation of retrospective cohorts was conducted. Patients were treated at home by the HaH units of three Spanish hospitals between January 2012 and December 2013. From the cost accounting of HaH OPAT (staff, pharmacy, transportation, diagnostic tests and structural), the cost of each outpatient course was obtained following a top-down strategy based on the use of resources. Costs associated with inpatient stay, if any, were estimated based on length of stay and ICD-9-CM diagnosis. There were 1324 HaH episodes in 1190 patients (median age 70 years). The median (interquartile range) stay at home was 10 days (7-15 days). Of the OPAT episodes, 91.5% resulted in cure or improvement on completion of intravenous therapy. The mean total cost of each infectious episode was €6707 [95% confidence interval (CI) €6189-7406]. The mean cost per OPAT episode was €1356 (95% CI €1247-1560), mainly distributed between healthcare staff costs (46%) and pharmacy costs (39%). The mean cost of inpatient hospitalisation of an infectious episode was €4357 (95% CI €3947-4977). The cost per day of inpatient hospitalisation was €519, whilst the cost per day of OPAT was €98, meaning a saving of 81%. This study shows that OPAT administered by HaH units resulted in lower costs compared with inpatient care in Spain.

Keywords: OPAT; Outpatient parenteral antimicrobial therapy; Intravenous antimicrobials; Hospital at home; Cost analysis

A.8 Article A8

Títol: An in-depth analysis shows a hidden atherogenic lipoprotein profile in non-diabetic chronic kidney disease patients

Autors: Bermudez-Lopez M, Forné C, Amigo N, Bozic M, Arroyo D, Bretones T, Alonso N, Cambray S, Del Pino MD, Mauricio D, Gorriz JL, Fernandez E, Valdivielso JM

Revista: Expert Opin Ther Targets. 2019 Jul;23(7):619-630

Factor d'impacte de l'any anterior de la publicació: 4,621 (percentil 88,6 a *Pharmacology & Pharmacy*)

DOI: 10.1080/14728222.2019.1620206

Resum original:

Background: Chronic kidney disease (CKD) is an independent risk factor for atherosclerotic disease. We hypothesized that CKD promotes a proatherogenic lipid profile modifying lipoprotein composition and particle number.

Methods: Cross-sectional study in 395 non-diabetic individuals (209 CKD patients and 186 controls) without statin therapy. Conventional lipid determinations were combined with advanced lipoprotein profiling by nuclear magnetic resonance, and their discrimination ability was assessed by machine learning.

Results: CKD patients showed an increase of very-low-density (VLDL) particles and a reduction of LDL particle size. Cholesterol and triglyceride content of VLDLs and intermediate-density (IDL) particles increased. However, low-density (LDL) and high-density (HDL) lipoproteins gained triglycerides and lost cholesterol. Total-Cholesterol, HDL-Cholesterol, LDL-Cholesterol, non-HDL-Cholesterol and Proprotein convertase subtilisin-kexin type (PCSK9) were negatively associated with CKD stages, whereas triglycerides, lipoprotein(a), remnant cholesterol, and the PCSK9/LDL-Cholesterol ratio were positively associated. PCSK9 was positively associated with total-Cholesterol, LDL-Cholesterol, LDL-triglycerides, LDL particle number, IDL-Cholesterol, and remnant cholesterol. Machine learning analysis by random forest revealed that new parameters have a higher discrimination ability to classify patients into the CKD group, compared to traditional parameters alone: area under the ROC curve (95% CI), .789 (.711, .853) vs .687 (.611, .755).

Conclusions: non-diabetic CKD patients have a hidden proatherogenic lipoprotein profile.

Keywords: Atherosclerosis, chronic kidney disease, dyslipidemia, lipoprotein subfractions, Lp(a), PCSK9

Bibliografía

1. Gisbert i Gelonch R. Economía y salud: economía, gestión económica y evaluación económica en el ámbito sanitario. Masson, 2002.
2. Robbins L. An essay on the nature and significance of economic science. Macmillan, 1932.
3. Drummond MF i McGuire A. Economic evaluation in health care: merging theory with practice. Oxford University Press, 2010.
4. Antonio Sacristán J, Ortún V, Rovira J, Prieto L i García-Alonso F. Evaluación económica en medicina. Med Clin (Barc) 2004;122:379-382.
5. Haines T, O'Brien L, McDermott F et al. A novel research design can aid disinvestment from existing health technologies with uncertain effectiveness, cost-effectiveness, and/or safety. J Clin Epidemiol 2014;67:144-151.
6. Niven DJ, Mrklas KJ, Holodinsky JK et al. Towards understanding the de-adoption of low-value clinical practices: a scoping review. BMC Med 2015;13:255.
7. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL i Torrance GW. Methods for the economic evaluation of health care programmes. Oxford University Press, 2015.
8. Briggs AH i O'Brien BJ. The death of cost-minimization analysis? Health Econ 2001;10:179-184.
9. Dakin H i Wordsworth S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. Health Econ 2013;22:22-34.
10. Pitt C, Goodman C i Hanson K. Economic evaluation in global perspective: a bibliometric analysis of the recent literature. Health Econ 2016;25:9-28.
11. Robinson R. Cost-utility analysis. BMJ 1993;307:859-862.
12. Prieto L, Sacristán JA, Pinto JL, Badia X, Antoñanzas F i Llano J del. Análisis de costes y resultados en la evaluación económica de las intervenciones sanitarias. Med Clin (Barc) 2004;122:423-429.
13. Janssen MF, Szende A, Cabases J, Ramos-Goñi JM, Vilagut G i König HH. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. Eur J Health Econ 2019;20:205-216.
14. Robinson R. Cost-benefit analysis. BMJ 1993;307:924-926.
15. Junoy J, Pinto J i Matarrodona E. El valor monetario de la salud. Economía de la salud y gestión sanitaria. Springer, 2001.
16. Puig-Junoy J, Pinto-Prades J i Ortún-Rubio V. El análisis coste-beneficio en sanidad. Aten Primaria 2001;27:422-427.

17. Johannesson M i Jönsson B. Economic evaluation in health care: Is there a role for cost-benefit analysis? *Health Policy* 1991;17:1-23.
18. O'Brien B. Cost-benefit analysis, willingness to pay. A: *Wiley StatsRef: Statistics reference online*. American Cancer Society, 2014.
19. O'Brien B i Gafni A. When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Med Decis Making* 1996;16. PMID: 8818128:288-299.
20. Bridges JF, Hauber AB, Marshall D et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health* 2011;14:403-413.
21. Schulz KF, Altman DG, Moher D i CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
22. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC i Vandenberg JP. The STrengthening the Reporting of OBServational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453-1457.
23. Collins G, Groves T, Marušić A et al. Enhancing the QUALity and Transparency Of Health Research. [Internet]. URL: <http://www.equator-network.org/> (cons. 29-08-2019).
24. Husereau D, Drummond M, Petrou S et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health* 2013;16:e1-e5.
25. National Institute for Clinical Excellence (NICE). Guide to the methods of technology appraisal 2013. [Internet]. NICE. 2013. URL: <http://nice.org.uk/process/pmg9> (cons. 03-07-2019).
26. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Guideline for economic evaluations in healthcare. [Internet]. IQWiG. 2017. URL: <https://www.iqwig.de/en/methods/methods-paper.3020.html> (cons. 03-07-2019).
27. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. 3rd ed. [Internet]. CADTH. Ottawa (ON), 2017. URL: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf (cons. 03-07-2019).
28. Pharmaceutical Benefits Advisory Committee (PBAC). Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (PBAC). [Internet]. PBAC. 2015. URL: <https://pbac.pbs.gov.au/information/printable-version-of-guidelines.html> (cons. 03-07-2019).
29. Puig-Junoy J, Oliva-Moreno J, Trapero-Bertrán M, Abellán-Perpiñán JM, Brosa-Riestra M i Servei Català de la Salut (CatSalut). Guia i recomanacions per a la realització i presentació d'avaluacions econòmiques i anàlisis d'impacte pressupostari de medicaments en l'àmbit del CatSalut. [Internet]. Generalitat de Catalunya. Departament de Salut. Servei Català de la Salut. Barcelona, 2014. URL: <https://catsalut>.

- gencat . cat / web / .content / minisite / catsalut / proveidors _ professionals / medicaments _ farmacia / farmaeconomica / caeip / gaeip _ publica _ catala _ octubre2014 _ catsalut . pdf (cons. 03-07-2019).
30. Zorginstituut Nederland. Guideline for economic evaluations in healthcare. [Internet]. 2016. URL: <https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare> (cons. 03-07-2019).
 31. Haute Autorité de santé (HAS). Department of Economics and Public Health Assessment. Choices in methods for economic evaluation. [Internet]. HAS. 2012. URL: https://www.has-sante.fr/portail/jcms/c_2035665/en/methods-for-health-economic-evaluation (cons. 03-07-2019).
 32. Oliva J, Brosa M, Espín J, Figueras M i Trapero M. Cuestiones controvertidas en evaluación económica (I): perspectiva y costes de intervenciones sanitarias. *Rev Esp Salud Publica* 2015;89:5-14.
 33. López Bastida J, Oliva J, Antoñanzas F et al. Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. *Gac Sanit* 2010;24:154-170.
 34. Weinstein MC, O'Brien B, Hornberger J et al. 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 Health* 2003;6:9-17.
 35. O'Mahony JF, Newall AT i Rosmalen J van. Dealing with time in health economic evaluation: eethodological issues and recommendations for practice. *Pharmacoeconomics* 2015;33:1255-1268.
 36. Prieto L, Sacristán JA, Antoñanzas F, Rubio-Terrés C, Pinto JL i Rovira J. Análisis coste-efectividad en la evaluación económica de intervenciones sanitarias. *Med Clin (Barc)* 2004;122:505-510.
 37. Smith R. New BMJ policy on economic evaluations: We won't publish economic evaluations unless offered the clinical results as well. *BMJ* 2002;325:1124.
 38. NHS Improvement and NHS England. National Tariff Payment System (NTPS). [Internet]. National Health System (NHS), 2019. URL: <https://improvement.nhs.uk/resources/national-tariff/> (cons. 04-07-2019).
 39. Medicare Benefits Schedule book. [Internet]. Australian Government. Department of Health, 2019. URL: <http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home> (cons. 04-07-2019).
 40. Mathes T, Jacobs E, Morfeld JC i Pieper D. Methods of international health technology assessment agencies for economic evaluations: a comparative analysis. *BMC Health Serv Res* 2013;13:371.
 41. Gol Montserrat J, García López JL, del Llano Señarís JE, Martín del Burgo M, Quecedo Gutiérrez L i Pi Corrales G. Análisis de costes indirectos: el caso del cáncer. Fundación Gaspar Casal, 2015.

42. Pike J i Grosse SD. Friction cost estimates of productivity costs in cost-of-illness studies in comparison with human capital estimates: a review. *Appl Health Econ Health Policy* 2018;16:765-778.
43. Trapero-Bertran M, Brosa Riestra M, Espín Balbino J i Oliva J. Cuestiones controvertidas en evaluación económica (II): medidas de resultado en salud de las intervenciones sanitarias. *Rev Esp Salud Publica* 2015;89:125-135.
44. Mehrez A i Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989;9:142-149.
45. Murray C, López A, World Health Organization, World Bank i Harvard School of Public Health. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. *Global burden of disease and injury series*. Harvard School of Public Health, 1996.
46. Al-Janabi H, Flynn TN i Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res* 2012;21:167-176.
47. Benson T. The Load Model: an alternative to QALY. *J Med Econ* 2017;20. PMID: 27559918:107-113.
48. Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Making* 1995;15:201-208.
49. Weinstein MC. A QALY is a QALY is a QALY — or is it? *J Health Econ* 1988;7:289-290.
50. Whitehead SJ i Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull* 2010;96:5-21.
51. International Society for Pharmacoeconomics and Outcomes Research. ISPOR Pharmacoeconomic Guidelines Around The World. [Internet]. URL: <https://tools.ispor.org/peguidelines/> (cons. 06-07-2019).
52. Revicki DA i Frank L. Pharmacoeconomic evaluation in the real world: effectiveness versus efficacy studies. *Pharmacoeconomics* 1999;15:423-434.
53. Frieden TR. Evidence for health decision making — Beyond randomized, controlled trials. *N Engl J Med* 2017;377. Ed. de Drazen JM, Harrington DP, McMurray JJ, Ware JH i Woodcock J:465-475.
54. Rubio-Terrés C, Sacristán JA, Badia X, Cobo E i García Alonso F. Métodos utilizados para realizar evaluaciones económicas de intervenciones sanitarias. *Med Clin (Barc)* 2004;122:578-583.
55. Ramsey SD, Willke RJ, Glick H et al. Cost-effectiveness analysis alongside clinical trials II: an ISPOR good research practices Task Force report. *Value Health* 2015;18:161-172.
56. Ramsey S, Willke R, Briggs A et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005;8:521-533.
57. Thorpe KE, Zwarenstein M, Oxman AD et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464-475.

58. The STRATOS Initiative. Causal inference. Topic Group 7 of the STRATOS Initiative. [Internet]. URL: http://www.stratos-initiative.org/group_7 (cons. 08-07-2019).
59. Hernán MA i Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;60:578-586.
60. Generalitat de Catalunya - Departament de Salut. Pla Director de Sistemes d'Informació del SISCAT. [Internet]. 2017. URL: <https://pdsis.blog.gencat.cat/> (cons. 29-07-2019).
61. Bernouilli D. Essai d'une nouvelle analyse de la mortalite causee pour la verole et des avantages de l'incubation pour la prevenir. *Mem Math Phys Acad Roy Sci (Paris)* 1760;1:1-45.
62. Hamer W. The Milroy Lectures on epidemic disease in England: the evidence of variability and of persistency of type. *Milroy lectures*. Bedford Press, 1906.
63. Ross R. The prevention of malaria. Dutton, 1910.
64. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW i McKenzie FE. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. Ed. de Chitnis CE. 2012.
65. Foppa I. A historical introduction to mathematical modeling of infectious diseases: seminal papers in epidemiology. Elsevier Science, 2016.
66. Younis T, Thana M i Skedgel C. Evidence in medicine: math versus biology! *Curr Oncol* 2017;24:349-351.
67. Egger M, Johnson L, Althaus C et al. Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. *F1000Res* 2018;6:1584.
68. Buxton MJ, Drummond MF, Van Hout BA et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;6:217-227.
69. Brosa M. La utilidad de la modelización clínico-económica en la investigación de resultados en salud. A: *La investigación de resultados en salud*. Ed. de Badia X. Edimac, 2000:232-233.
70. Caro JJ, Briggs AH, Siebert U, Kuntz KM i ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Value Health* 2012;15:796-803.
71. Barton P, Bryan S i Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004;9:110-118.
72. Brennan A, Chick SE i Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15:1295-1310.
73. Cooper K, Brailsford SC i Davies R. Choice of modelling technique for evaluating health care interventions. *J Oper Res Soc* 2007;58:168-176.
74. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008;26:131-148.

75. Roberts M, Russell LB, Paltiel AD et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Value Health* 2012;15:804-811.
76. Tsoi B, O'Reilly D, Jegathisawaran J, Tarride JE, Blackhouse G i Goeree R. Systematic narrative review of decision frameworks to select the appropriate modelling approaches for Health Economic evaluations. *BMC Res Notes* 2015;8.
77. Tsoi B, Goeree R, Jegathisawaran J, Tarride JE, Blackhouse G i O'Reilly D. Do different decision-analytic modeling approaches produce different results? A systematic review of cross-validation studies. *Expert Rev Pharmacoecon Outcomes Res* 2015;15:451-463.
78. Siebert U, Alagoz O, Bayoumi AM et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value Health* 2012;15:812-820.
79. Karnon J, Stahl J, Brennan A et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value Health* 2012;15:821-827.
80. Caro JJ i Möller J. Advantages and disadvantages of discrete-event simulation for health economic analyses. *Expert Rev Pharmacoecon Outcomes Res* 2016;16:327-329.
81. Pitman R, Fisman D, Zaric GS et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. *Value Health* 2012;15:828-834.
82. Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health* 2012;15:843-850.
83. Rubio-Terrés C, Cobo E, Antonio Sacristán J, Prieto L, Llano J del i Badia X. Análisis de la incertidumbre en las evaluaciones económicas de intervenciones sanitarias. *Med Clin (Barc)* 2004;122:668-674.
84. Drummond M, Brown R, Fendrick AM et al. Use of pharmacoeconomics information—report of the ISPOR Task Force on use of pharmacoeconomic/Health Economic information in health-care decision making. *Value Health* 2003;6:407-416.
85. Briggs AH i Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;3:1-134.
86. Briggs AH, Weinstein MC, Fenwick EA et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health* 2012;15:835-842.
87. Polsky D, Glick HA, Willke R i Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ* 1997;6:243-252.
88. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijnenbrok-Kal M i Hunink MM. Uncertainty and patient heterogeneity in medical decision models. *Med Decis Making* 2010;30:194-205.

89. Griffin S, Claxton K, Hawkins N i Sculpher M. Probabilistic analysis and computationally expensive models: necessary and required? *Value Health* 2006;9:244-252.
90. Carvalho TM de, Heijnsdijk EAM, Coffeng L i Koning HJ de. Evaluating parameter uncertainty in a simulation model of cancer using emulators. *Med Decis Making* 2019;39:405-413.
91. O'Hagan A, Buck CE, Daneshkhah A et al. *Uncertain judgements: eliciting experts' probabilities*. Chichester, UK: John Wiley & Sons, Ltd, 2006.
92. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;26:781-798.
93. Claxton K, Sculpher M i Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). *Lancet* 2002;360:711-715.
94. Claxton K, Neumann PJ, Araki S i Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. *Int J Technol Assess Health Care* 2001;17:38-55.
95. Johannesson M i Weinstein MC. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993;12:459-467.
96. Drummond M, Torrance G i Mason J. Cost-effectiveness league tables: more harm than good? *Soc Sci Med* 1993;37:33-40.
97. McCullagh P(i Nelder JA. *Generalized linear models*. Chapman i Hall, 1989:511.
98. Hosmer D, Lemeshow S i Sturdivant R. *Applied logistic regression*. Wiley Series in Probability and Statistics. Wiley, 2013.
99. Tang W, He H i Tu XM. *Applied categorical and count data analysis*. CRC Press, 2012:363.
100. Therneau TM i Grambsch PM. *Modeling survival data: extending the Cox model*. Statistics for Biology and Health. New York, NY: Springer New York, 2000.
101. Collett D. *Modelling survival data in medical research*. 3a ed. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2015.
102. Putter H, Fiocco M i Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007;26:2389-2430.
103. Willan AR i Briggs AH. *Statistical analysis of cost-effectiveness data*. John Wiley, 2006:196.
104. Hernán MA, Clayton D i Keiding N. The Simpson's paradox unraveled. *Int J Epidemiol* 2011;40:780-785.
105. Mihaylova B, Briggs A, O'Hagan A i Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ* 2011;20:897-916.
106. Glick HA, Doshi JA, Sonnad SS i Polsky D. *Economic evaluation in clinical trials*. Oxford University Press, 2014.
107. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC Biostatistics Series. Taylor & Francis, 2012.

108. Liu L, Wolfe RA i Kalbfleisch JD. A shared random effects model for censored medical costs and mortality. *Stat Med* 2007;26:139-155.
109. Liu L. Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Stat Med* 2009;28:972-986.
110. Xu H, Daggy J, Yu D, Craig BA i Sands L. Joint modeling of medical cost and survival in complex sample surveys. *Stat Med* 2013;32:1509-1523.
111. Piulachs X, Alemany R i Guillen M. Joint modelling of survival and emergency medical care usage in spanish insureds aged 65+. *PLoS One* 2016;11. Ed. de Lu T:e0153234.
112. Yu KH, Beam AL i Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;2:719-731.
113. Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health* 2015;18:137-140.
114. Doupe P, Faghmous J i Basu S. Machine learning for health services researchers. *Value Health* 2019;22:808-815.
115. Borenstein M, Hedges L, Higgins J i Rothstein H. *Introduction to meta-analysis*. Wiley, 2011.
116. Catalá-López F, Tobías A i Roqué M. Conceptos básicos del metaanálisis en red. *Aten Primaria* 2014;46:573-581.
117. European Cancer Information System. Cancer burden statistics and trends across Europe. [Internet]. URL: <https://ecis.jrc.ec.europa.eu/> (cons. 30-08-2019).
118. Timmis A, Townsend N, Gale C et al. European Society of Cardiology: cardiovascular disease statistics 2017. *Eur Heart J* 2018;39:508-579.
119. Luengo-Fernandez R, Gray AM i Rothwell PM. A population-based study of hospital care costs during 5 years after transient ischemic attack and stroke. *Stroke* 2012;43:3343-3351.
120. Walker S, Asaria M, Manca A et al. Long-term healthcare use and costs in patients with stable coronary artery disease: a population-based cohort using linked health records (CALIBER). *Eur Heart J Qual Care Clin Outcomes* 2016;2:125-140.
121. Mozaffarian D, Benjamin EJ, Go AS et al. Heart disease and stroke statistics—2016 update. *Circulation* 2016;133.
122. Townsend N, Wilson L, Bhatnagar P, Wickramasinghe K, Rayner M i Nichols M. Cardiovascular disease in Europe: epidemiological update 2016. *Eur Heart J* 2016;37:3232-3245.
123. Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, Burns R, Rayner M, Townsend N. European cardiovascular disease statistics 2017. [Internet]. European Heart Network. Brussels, 2017. URL: <http://www.ehnheart.org/cvd-statistics.html> (cons. 30-08-2019).
124. Nicholson G, Gandra S, Halbert R, Richarriya A i Nordyke R. Patient-level costs of major cardiovascular conditions: a review of the international literature. *Clinicoecon Outcomes Res* 2016;Volume 8:495-506.

125. Ryder S, Fox K, Rane P et al. A systematic review of direct cardiovascular event costs: an international perspective. *Pharmacoeconomics* 2019;37:895-919.
126. Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Collaborators. The impact of mammography and adjuvant therapy on US breast cancer mortality (1975-2000): collective results from the Cancer Intervention and Surveillance Modeling Network. *J Natl Cancer Inst Monographs* 2006;36:1-126.
127. Berry DA, Cronin KA, Plevritis SK et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;353. PMID: 16251534:1784-1792.
128. Román R, Sala M, Salas D et al. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol* 2011;23:104-111.
129. Kalager M, Adami HO, Bretthauer M i Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian Screening Program. *Ann Intern Med* 2012;156:491-499.
130. Jørgensen KJ i Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339.
131. Mandelblatt JS, Cronin KA, Bailey S et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009;151:738-747.
132. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R i von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition—summary document. *Ann Oncol* 2008;19:614-622.
133. Institute of Medicine and National Research Council. Saving women's lives: strategies for improving breast cancer detection and diagnosis. Ed. de Joy JE, Penhoet EE i Petitti DB. Washington, DC: The National Academies Press, 2005.
134. Arrospide A, Forné C, Rué M, Torà N, Mar J i Baré M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. *BMC Cancer* 2013;13:587.
135. Barton P, Andronis L, Briggs A, McPherson K i Capewell S. Effectiveness and cost effectiveness of cardiovascular disease prevention in whole populations: modelling study. *BMJ* 2011;343:d4044.
136. Kypridemos C, Allen K, Hickey GL et al. Cardiovascular screening to reduce the burden from cardiovascular disease: microsimulation study to quantify policy options. *BMJ* 2016;i2793.
137. Mytton OT, Jackson C, Steinacher A et al. The current and potential health benefits of the National Health Service Health Check cardiovascular disease prevention programme in England: a microsimulation study. *PLoS Med* 2018;15. Ed. de Sheikh A:e1002517.
138. Kypridemos C, Collins B, McHale P et al. Future cost-effectiveness and equity of the NHS Health Check cardiovascular disease prevention

- programme: microsimulation modelling using data from Liverpool, UK. *PLoS Med* 2018;15. Ed. de Sheikh A:e1002573.
139. Zhang QL i Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health* 2008;8:117.
 140. Webster AC, Nagler EV, Morton RL i Masson P. Chronic kidney disease. *Lancet* 2017;389:1238-1252.
 141. Go AS, Chertow GM, Fan D, McCulloch CE i Hsu Cy. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004;351:1296-1305.
 142. Tonelli M, Wiebe N, Cullerton B et al. Chronic kidney disease and mortality risk: a systematic review. *J Am Soc Nephrol* 2006;17:2034-2047.
 143. Weiner DE, Tighiouart H, Elsayed EF et al. The Framingham predictive instrument in chronic kidney disease. *J Am Coll Cardiol* 2007;50:217-224.
 144. Rosengren A, Wallentin L, Simoons M et al. Age, clinical presentation, and outcome of acute coronary syndromes in the Euroheart acute coronary syndrome survey. *Eur Heart J* 2006;27:789-795.
 145. Benjamin EJ, Blaha MJ, Chiuve SE et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation* 2017;135:e146-e603.
 146. Ibanez B, James S, Agewall S et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2018;39:119-177.
 147. Neumann FJ, Sousa-Uva M, Ahlsson A et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J* 2019;40:87-165.
 148. Garg A, Garg L, Agarwal M et al. Routine invasive versus selective invasive strategy in elderly patients older than 75 years with non-ST-segment elevation acute coronary syndrome: a systematic review and meta-analysis. *Mayo Clin Proc* 2018;93:436-444.
 149. Schoenenberger AW, Radovanovic D, Windecker S et al. Temporal trends in the treatment and outcomes of elderly patients with acute coronary syndrome. *Eur Heart J* 2016;37:1304-1311.
 150. Malta Hansen C, Wang TY, Chen AY et al. Contemporary patterns of early coronary angiography use in patients with non-ST-segment elevation myocardial infarction in the United States: insights from the National Cardiovascular Data Registry Acute Coronary Treatment and Intervention Outcomes Network Registry. *JACC Cardiovasc Interv* 2018;11:369-380.
 151. Madhavan MV, Gersh BJ, Alexander KP, Granger CB i Stone GW. Coronary artery disease in patients ≥ 80 years of age. *J Am Coll Cardiol* 2018;71:2015-2040.
 152. Fox KAA, Clayton TC, Damman P et al. Long-term outcome of a routine versus selective invasive strategy in patients with non-ST-segment

- elevation acute coronary syndrome a meta-analysis of individual patient data. *J Am Coll Cardiol* 2010;55:2435-2445.
153. Valgimigli M, Bueno H, Byrne RA et al. 2017 ESC focused update on dual antiplatelet therapy in coronary artery disease developed in collaboration with EACTS: the Task Force for dual antiplatelet therapy in coronary artery disease of the European Society of Cardiology (ESC) and of the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2018;39:213-260.
 154. Skolnick AH, Alexander KP, Chen AY et al. Characteristics, management, and outcomes of 5,557 patients age ≥ 90 years with acute coronary syndromes: results from the CRUSADE Initiative. *J Am Coll Cardiol* 2007;49:1790-1797.
 155. Ndrepepa G, Neumann FJ, Schulz S et al. Incidence and prognostic value of bleeding after percutaneous coronary intervention in patients older than 75 years of age. *Catheter Cardiovasc Interv* 2014;83:182-189.
 156. Dégano IR, Subirana I, Fusco D et al. Percutaneous coronary intervention reduces mortality in myocardial infarction patients with comorbidities: implications for elderly patients with diabetes or kidney disease. *Int J Cardiol* 2017;249:83-89.
 157. Schousboe JT, Kerlikowske K, Loh A i Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med* 2011;155:10.
 158. Ravesteyn NT van, Miglioretti DL, Stout NK et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. *Ann Intern Med* 2012;156:609-617.
 159. Stout NK, Lee SJ, Schechter CB et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. *J Natl Cancer Inst* 2014;106:dju092.
 160. Sprague BL, Stout NK, Schechter C et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. *Ann Intern Med* 2015;162:157.
 161. Trentham-Dietz A, Kerlikowske K, Stout NK et al. Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or older: collaborative modeling of screening outcomes. *Ann Intern Med* 2016;165:700.
 162. Pashayan N, Morris S, Gilbert FJ i Pharoah PDP. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncol* 2018;4:1504-1510.
 163. Tosteson AN, Stout NK, Fryback DG et al. Cost-effectiveness of digital mammography breast cancer screening. *Ann Intern Med* 2008;148:1.
 164. Ayer T, Alagoz O i Stout NK. A POMDP approach to personalize mammography screening decisions. *Oper Res* 2012;60:1019-1034.
 165. Arnold M. Simulation modeling for stratified breast cancer screening – a systematic review of cost and quality of life assumptions. *BMC Health Serv Res* 2017;17:802.

166. Gail MH, Brinton LA, Byar DP et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879-1886.
167. Chen J, Pee D, Ayyagari R et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J Natl Cancer Inst* 2006;98:1215-1226.
168. Barlow WE, White E, Ballard-Barbash R et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 2006;98:1204-1214.
169. Tice JA, Cummings SR, Ziv E i Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat* 2005;94:115-122.
170. Vachon CM, Brandt KR, Ghosh K et al. Mammographic breast density as a general marker of breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2007;16:43-49.
171. Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R i Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. *J Natl Cancer Inst* 2013;105:1043-1049.
172. Mavaddat N, Pharoah PDP, Michailidou K et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 2015;107.
173. Vachon CM, Pankratz VS, Scott CG et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst* 2015;107.
174. Shieh Y, Hu D, Ma L et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat* 2016;159:513-525.
175. Louro J, Posso M, Hilton Boon M et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer* 2019;121:76-85.
176. Kerlikowske K, Ichikawa L, Miglioretti DL et al. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. *J Natl Cancer Inst* 2007;99:386-395.
177. Lokate M, Stellato RK, Veldhuis WB, Peeters PHM i Gils CH van. Age-related changes in mammographic density and breast cancer risk. *Am J Epidemiol* 2013;178:101-109.
178. Work ME, Reimers LL, Quante AS, Crew KD, Whiffen A i Terry MB. Changes in mammographic density over time in breast cancer cases and women at high risk for breast cancer. *Int J Cancer* 2014;135:1740-1744.
179. Kerlikowske K, Gard CC, Sprague BL, Tice JA, Miglioretti DL i Breast Cancer Surveillance Consortium. One versus two breast density measures to predict 5- and 10-year breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2015;24:889-897.
180. Major RW, Cheng MRI, Grant RA et al. Cardiovascular disease risk factors in chronic kidney disease: a systematic review and meta-analysis. *PLoS One* 2018;13. Ed. de Reboldi G:e0192895.

181. Barreto DV, Lenglet A, Liabeuf S et al. Prognostic implication of plasma osteopontin levels in patients with chronic kidney disease. *Nephron Clin Pract* 2011;117:363-372.
182. Yilmaz MI, Siriolpol D, Saglam M et al. Osteoprotegerin in chronic kidney disease: associations with vascular damage and cardiovascular events. *Calcif Tissue Int* 2016;99:121-130.
183. Yabluchanskiy A, Ma Y, Iyer RP, Hall ME i Lindsey ML. Matrix metalloproteinase-9: many shades of function in cardiovascular disease. *Physiology (Bethesda)* 2013;28:391-403.
184. Holm PW, Slart RHJA, Zeebregts CJ, Hillebrands JL i Tio RA. Atherosclerotic plaque development and instability: a dual role for VEGF. *Ann Med* 2009;41:257-264.
185. Sanchis J, Núñez J, Bodí V et al. Influence of Comorbid Conditions on One-Year Outcomes in Non-ST-Segment Elevation Acute Coronary Syndrome. *Mayo Clin Proc* 2011;86:291-296.
186. Fox KAA, Anderson FA, Dabbous OH et al. Intervention in acute coronary syndromes: do patients undergo intervention on the basis of their risk characteristics? The Global Registry of Acute Coronary Events (GRACE). *Heart* 2007;93:177-182.
187. Bebb O, Smith FG, Clegg A, Hall M i Gale CP. Frailty and acute coronary syndrome: A structured literature review. *Eur Heart J Acute Cardiovasc Care* 2018;7:166-175.
188. Rue M, Vilaprinyo E, Lee S et al. Effectiveness of early detection on breast cancer mortality reduction in Catalonia (Spain). *BMC Cancer* 2009;9:326.
189. Carles M, Vilaprinyo E, Cots F et al. Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). *BMC Cancer* 2011;11:192.
190. Gøtzsche PC i Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013:CD001877.
191. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *The Lancet* 2012;380:1778-1786.
192. Paci E i EUROSCREEN Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19 Suppl 1:5-13.
193. Domingo L, Sala M, Servitja S et al. Phenotypic characterization and risk factors for interval breast cancers in a population-based breast cancer screening program in Barcelona, Spain. *Cancer Causes Control* 2010;21:1155-1164.
194. McCann J, Britton PD, Warren RM i Hunnam G. Radiological peer review of interval cancers in the East Anglian breast screening programme: what are we missing? *East Anglian Breast Screening Programme. J Med Screen* 2001;8:77-85.
195. Kirsh VA, Chiarelli AM, Edwards SA et al. Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program. *J Natl Cancer Inst* 2011;103:942-950.

196. García-Blas S, Bonanad C i Sanchis J. Invasive strategy in elderly patients with acute coronary syndrome in 2018: close to the truth? *J Geriatr Cardiol* 2019;16:114-120.
197. André R, Bongard V, Elosua R et al. International differences in acute coronary syndrome patients' baseline characteristics, clinical management and outcomes in Western Europe: the EURHOBOP study. *Heart* 2014;100:1201-1207.
198. Everett BM i Ridker PM. Biomarkers for cardiovascular screening: progress or passé? *Clin Chem* 2017;63:248-251.
199. Drucker E i Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J* 2013;4:7.
200. Shieh Y, Eklund M, Madlensky L et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst* 2017;109:djw290.
201. Esserman LJ i WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* 2017;3:34.
202. Delalogue S. My Personalized Breast Screening (MyPeBS). A: *ClinicalTrials.gov [Internet]. Bethesda (MD): US National Library of Medicine 2000. ClinicalTrials.gov Identifier: NCT03672331. 14 set 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03672331 (cons. 17-04-2020). [darrera actualització: 16 abr. 2020].*
203. Mantellini P. Tailored Screening for Breast Cancer in Premenopausal Women (TBST). A: *ClinicalTrials.gov [Internet]. Bethesda (MD): US National Library of Medicine 2000. ClinicalTrials.gov Identifier: NCT02619123. 2 des 2015. URL: https://clinicaltrials.gov/ct2/show/NCT02619123 (cons. 17-04-2020). [darrera actualització: 10 febr. 2017].*
204. Bolívar B, Fina Avilés F, Morros R et al. Base de datos SIDIAP: la historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Med Clin (Barc)* 2012;138:617-621.
205. Ramos R, Balló E, Marrugat J et al. Validez del Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP) en el estudio de enfermedades vasculares: estudio EMMA. *Rev Esp Cardiol* 2012;65:29-37.
206. Carles M, Martínez-Alonso M, Pons A et al. The effect of information about the benefits and harms of mammography on women's decision-making: study protocol for a randomized controlled trial. *Trials* 2017;18:426.
207. Pérez-Lacasta MJ, Martínez-Alonso M, Garcia M et al. Effect of information about the benefits and harms of mammography on women's decision making: the InforMa randomised controlled trial. *PLoS One* 2019;14. Ed. d'Houston A:e0214057.

208. Rué M. Assessment of Feasibility and Acceptability of Personalized Breast Cancer Screening (DECIDO). A: *ClinicalTrials.gov [Internet]. Bethesda (MD): US National Library of Medicine 2000. ClinicalTrials.gov Identifier: NCT03791008. 1 gen 2019. URL: <https://clinicaltrials.gov/ct2/show/NCT03791008> (cons. 17-04-2020). [darrera actualització: 7 febr. 2019].*
209. Adamson AS i Welch HG. Machine learning and the cancer-diagnosis problem — No gold standard. *N Engl J Med* 2019;381:2285-2287.