IMPACTO EPIDEMIOLÓGICO Y ECONÓMICO DEL PROGRAMA DE DETECCIÓN PRECOZ DEL CÁNCER DE MAMA EN EL PAÍS VASCO: Desarrollo de un modelo de Simulación

EPIDEMIOLOGICAL AND ECONOMIC IMPACT OF THE BREAST CANCER EARLY DETECTION PROGRAMME IN THE BASQUE COUNTRY: Development of a simulation model

Arantzazu Arrospide Elgarresta
UPV-EHU, Vitoria-Gasteiz, 2016

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

IMPACTO EPIDEMIOLÓGICO Y ECONÓMICO DEL PROGRAMA DE DETECCIÓN PRECOZ DEL CÁNCER DE MAMA EN EL PAÍS VASCO: Desarrollo de un modelo de Simulación.

Arantzazu Arrospide Elgarresta

UPV-EHU, Vitoria-Gasteiz, 2016

Tesis doctoral dirigida por:

Prof. Luis Carlos Abecia Inchaurregui

Prof. Javier Mar Medina

EPIDEMIOLOGICAL AND ECONOMIC IMPACT OF THE BREAST CANCER EARLY DETECTION PROGRAMME IN THE BASQUE COUNTRY: Development of a simulation model.

Arantzazu Arrospide Elgarresta

UPV-EHU, Vitoria-Gasteiz, 2016

Doctoral thesis supervised by:

Prof. Luis Carlos Abecia Inchaurregui

Prof. Javier Mar Medina

“Aitarena den txapel hau, eskeintzen diot amari,

haurra nintzela euskaraz erakutsi zidanari.”

Amets Arzallus

“Izar artean zehar hantxe ibiliko da,

irribarre eginez begira ezazu gora,

bestela bera ere triste jarriko da.”

Osabaren oroimenez

Índice

Abstract

INTRODUCTION: Since 1996 screening mammograms have been done in a biennial basis to women aged 50 to 69 in the Basque Country (Spain). Based on epidemiological observations and simulation techniques it is possible to extend observed short term data into anticipated long term results. In addition, such a mass preventive intervention must be evaluated also in economic terms to warrant that the allocated resources are a worthwhile investment for the entire population. Although currently, age and gender are the only criteria for defining the target population, a reliable individual BC risk estimate based on known risk factors could be helpful in order to personalize screening programs.

OBJECTIVES: The main objective of this study was the evaluation of the screening programme in terms of health and costs, in the Basque women population, since 1996 through 2011 including long term effects. Three individualized breast cancer risk prediction models were assessed as possible screening optimization tools.

METHODS: A discrete event simulation model was built to represent the natural history of breast cancer in women invited to the breast cancer screening programme in the Basque Country. The disease progress was described in three main states: healthy, preclinical and clinical. We made the following assumptions: 1) All women would be diagnosed at the beginning of the clinical phase unless they were diagnosed previously through the screening programme; 2) The survival distribution for clinically detected or screen-detected breast cancer only depends on the disease-stage at diagnosis; 3) Screening produces a stage-shift at diagnosis, with a more favorable distribution for screen-detected cases. The data collected during the 15 years when the screening programme was held allowed validating the model.

Both short and long term screening effects were measured as differences in benefits and harms between the screened and unscreened populations. Breast cancer mortality reduction and life-years gained were considered as screening benefits, whereas, overdiagnosis and false positive results were assessed as harms. Breast cancer costs included screening, diagnosis and treatment costs and effectiveness was assessed based on quality-adjusted life years. A sample of first-time participant women was used to obtain projections of risk in three and five years using the Gail, Chen and Barlow models.

RESULTS: During the evaluated period 414,041 women were invited to the screening programme and the mean attendance rate was 78%. Due to the screening programme 5,267 cancers were early detected. The screening programme yielded a 16% reduction in breast cancer mortality and a 10% increase in the incidence of breast cancer through 2011. However, analyzing a single cohort with lifetime follow-up, 4% of the screen-detected cancers were overdiagnosed. Survival time increased in 2.5 years of life for each screening-detected woman.

All the mammograms carried out during the evaluated years costed 55.3 million Euros. The extra cost induced by additional diagnosis tests for the positive mammograms in the screening programme was 12 million Euros. Although early diagnosis allowed saving 39 million Euros in breast cancer treatment during the studied period, in terms of total costs the background scenario had a lower cost on average. All the simulations resulted in an incremental cost-effectiveness ratio lower than 10,000 € per quality adjusted life year.

The Gail and Chen models showed good calibration while the Barlow model overestimated the number of cases. Although they passed the calibration test, the Gail and Chen models overestimated the number of cases in some breast density categories. The 5-year projection for the Chen and Barlow models had the highest discrimination, although it was not considered enough for their application in screening programmes.

CONCLUSIONS: Fifteen years after the screening programme started, this study supports an important decrease in breast cancer mortality, with reasonable risk of harm with screening. These pidemiological benefits related to the centralised screening system were confirmed by the economic results and sustain the continuation of the breast cancer screening programme in the Basque population. However, there is a need to optimize screening programmes taking into account individual risk. In this line, the studied risk models cannot be used as a measure of individual risk in early detection programs to customize screening strategies.

Resúmen

INTRODUCCIÓN: El cáncer de mama es el tumor maligno más frecuente entre mujeres de la Comunidad Autónoma del País Vasco y su incidencia ha aumentado significativamente desde la introducción del cribado poblacional el año 1996. Sin embargo la mortalidad asociada al cáncer de mama desciende un 2.8% anual desde el año 1992.

El cribado es un programa de salud pública en el que los miembros de una población definida, que no necesariamente perciben tener un mayor riesgo, son invitados a someterse a pruebas diagnósticas para reducir el riesgo de la enfermedad o sus complicaciones. Los factores de riesgo de cáncer de mama conocidos actualmente no son modificables por lo que no caben intervenciones poblacionales de prevención primaria. La decisión de implantar el programa poblacional de detección precoz de cáncer de mama fue tomada, en la mayoría de países desarrollados, en base a ensayos clínicos aleatorizados realizados entre 1963 y 1991.

En el País Vasco el Departamento de Sanidad del Gobierno Vasco puso en marcha el Programa de Detección Precoz del Cáncer de Mama en el año 1995 en el Área Sanitaria de Araba, extendiéndose a lo largo de 1997 a toda la CAPV. El test diagnostico del cribado consistió en la realización de una mamografía bilateral en doble proyección: craneo-caudal y oblicua medio-lateral con periodicidad bienal a las mujeres entre 50 y 64 anos y que se extendió hasta los 69 años a partir de 2006 de forma progresiva, completándose la ampliación hasta los 69 años, en el año 2010.

El alto coste de las intervenciones de salud pública por su gran población diana y la complejidad de su gestión hacen indispensable su evaluación en términos de recursos económicos, materiales y humanos. Para ello se requiere la definición de un marco estándar para la evaluación de intervenciones de salud pública que permita identificar el programa que más beneficio aporta a la población en relación a la inversión necesaria. El beneficio de las decisiones clínicas, es valorado desde el punto de vista de los pacientes por lo que el beneficio está directamente relacionado con las preferencias individuales. En la evaluación de programas de salud pública, sin embargo los beneficios y costes deben adoptar una perspectiva social, no individual.

Mediante técnicas de simulación y basándose en datos epidemiológicos observados a corto plazo es posible predecir los resultados a largo plazo. Este tipo de intervenciones preventivas debe además ser evaluado en términos económicos para garantizar el beneficio poblacional de la inversión realizada. La estimación del riesgo individual de cáncer de mama en base a factores de riesgo podría ayudar a la personalizar los programas de cribado.

OBJETIVOS: El principal objetivo fue la evaluación del programa de detección precoz del cáncer de mama entre 1996 y 2011 en términos de salud y costes en la población vasca de mujeres, incluyendo los efectos a largo plazo. Tres modelos de predicción del riesgo individual fueron evaluados como posibles herramientas para la optimización del programa de cribado.

MÉTODOS: Con el objetivo de estimar el impacto a largo plazo del programa, se construyó un modelo de simulación con eventos discretos para reproducir la historia natural del cáncer de mama. La evaluación incluyó a todas las mujeres invitadas al programa de detección precoz de cáncer de mama en el País Vasco desde su inicio en 1996 hasta el 31 de diciembre del 2011, dado que en el año 2012 el programa fue extendido para incluir a las mujeres con antecedentes familiares de primer grado a partir de los 40 años. Los datos registrados en el programa de cribado permitieron reproducir exactamente el número de mujeres invitadas por primera vez al programa cada año, así como su distribución de edad y la tasa de participación correspondientes.

En la historia natural del cáncer de mama se distinguieron tres estados de la enfermedad: sano, preclínico y clínico. El estado preclínico se define como el período asintomático durante el cual el cáncer puede ser detectado mediante cribado. La función de inicio del período preclínico se obtuvo de un estudio previo realizado en base a datos de Registro de Cáncer Catalán. Se aplicaron modelos lineales generalizados con distribución de Poisson para la parametrización de la incidencia de cáncer de mama en función de edad y la cohorte de nacimiento. En cuanto a la duración de la fase preclínica, los resultados de ensayos clínicos establecieron que sigue una distribución exponencial con 4 años de media.

Se asumió que todas las mujeres serían diagnósticadas al inicio de la fase clínica a no ser que fueran detectadas previamente mediante cribado. Por tanto, se utilizaron los datos del registro poblacional del País Vasco en el año 1995, antes de que se iniciara el programa de cribado, para establecer la distribución de los estadios de detección clínica. El avance en el diagnóstico debido al cribado resultó en una distribución más temprana de los estadios de detección. Los carcinomas in situ fueron consideraros el estadio más bajo en el que el cáncer podría ser detectado y fueron incluidos en el modelo. En los cánceres diagnósticados, tanto por la vía de cribado como en la fase clínica, la supervivencia dependió únicamente del estadio de detección del cáncer. Se utilizaron las funciones de supervivencia dependientes de la edad y el estadio de detección descritas por Vilaprinyó et al.

La mortalidad por causas distintas al cáncer de mama también fue incluida en el modelo como riesgo competitivo con la mortalidad por cáncer de mama. La edad de fallecimiento final resultó en el mínimo entre la edad de muerte por otras causas y la edad de muerte por cáncer de mama, asignadas aleatoria e independiente. Los datos de mortalidad general y específica en el período 1986-2010 se obtuvieron a partir del registro de mortalidad del País Vasco.

Fueron tres los parámetros no observables calibrados para reproducir los datos observados en los primeros 15 años del programa: el intervalo entre invitaciones sucesivas al cribado, la distribución de edad de inicio del período preclínico y la media de la distribución de la duración del estado preclínico. Se aplicó el estadístico Chi-cuadrado para la selección del valor óptimo de estos parámetros. La validación final del modelo se realizó mediante la comparación de los resultados del modelo poblacional en el escenario de cribado y los indicadores observados en el programa a lo largo del período evaluado.

Los efectos del cribado a corto y largo plazo se midieron comparando los resultados en la población cribada y la no cribada. La reducción en la mortalidad por cáncer de mama y los años de vida ganados fueron considerados beneficios del programa. Como daños se midieron el sobrediagnóstico y los resultados falsos positivos.

En el cálculo de los costes unitarios se incluyeron los costes sanitarios del diagnóstico y tratamientos distinguiendo tratamiento inicial y seguimiento. La técnica aplicada fue el método de micro-costes a partir de las guías de práctica clínica. Para ello se analizaron el consumo de recursos y los costes unitarios del Sistema Vasco de Salud.

Dada la falta de valores propios de calidad de vida para los diferentes estados de salud en el desarrollo del cáncer de mama, las utilidades aplicadas en el modelo fueron estimadas a partir de los valores de la población general Española en función de la edad. Se asumió a lo largo del tratamiento inicial del cáncer una reducción del 90% de la calidad de vida en los carcinomas In Situ o cánceres detectados en estadio I, del 75 % para los estadios II y III y del 60% el los cánceres metastásicos.

Mediante este modelo se calcularon los costes totales relativos al cáncer de mama (cribado, diagnóstico y tratamiento) y los años de vida ajustados por calidad para toda la vida de la población diana en escenario con y sin cribado. El diseño probabilístico del modelo permitió variar aleatoriamente los parámetros fundamentales del modelo. Se utilizó la distribución Uniforme para el tiempo entre invitaciones y el tiempo medio en estado pre-clínico, la distribución Beta para la sensibilidad y la especificidad y Dirichlet para la distribución de los estadios de detección en los cánceres detectados mediante el cribado.

Así mismo el modelo permitió analizar el impacto presupuestario del programa de cribado de cáncer de mama en el País Vasco. El modelo poblacional calculó el coste anual del cáncer de mama incluyendo el cribado, diagnóstico y tratamientos para la población de mujeres invitadas al programa de cribado en el escenario con y sin cribado.

Finalmente, se utilizó una muestra de 13760 mujeres que participaban por primera vez en el cribado de Sabadell-Cerdanyola entre Octubre de 1995 y Junio de 1998 para obtener las proyecciones del riesgo individual a tres y cinco años utilizando los modelos de Gail, Chen y Barlow. Puesto que estos modelos fueron validados sin incluir los carcinomas in situ tampoco fueron incluidos como cánceres en esta evaluación.

El modelo de Gail incluyó como factores de riesgo el número de familiares de primer grado con antecedentes de cáncer de mama, la edad en el primer parto, la edad de

menarquia y el número de biopsias previas realizadas. El modelo de Chen, en cambio, no incluyó la edad de menarquia ni las interacciones, sin embargo, sí tuvo en cuenta la densidad mamaria y el peso de la mujer. Estos modelos con estructura idéntica permiten ajustar la incidencia de cáncer de mama y la mortalidad a la población a estudio y calcular el riesgo de desarrollar cáncer de mama en cualquier periodo establecido.

El modelo de Barlow en cambió fue diseñado para estimar el riesgo de desarrollar cáncer a lo largo de un año. Siguiendo la recomendación del artículo original se asumió que la probabilidad de desarrollar cáncer de mama en años sucesivos era idéntica e independiente para calcular el riesgo para tres y cinco años. El modelo de Barlow incluyó además de la densidad mamaria, el uso de la terapia hormonal sustitutiva, el índice de masa corporal, el resultado de mamografías previas realizadas, la raza y la etnia.

Se utilizó el estadístico de ajuste C de Hosmer-Lemeshow para medir la calibración de los modelos a tres y cinco años. La discriminación de los modelos fue evaluado mediante el estadístico C de Harrell y el área bajo la curva Receiver Operating Characteristic.

RESULTADOS: Durante el periodo evaluado 414,041 mujeres fueron invitadas a participar en el programa con una tasa de participación del 78%. Se detectaron 5,267 canceres mediante cribado. El programa de cribado consiguió una reducción del 16% en la mortalidad por cáncer de mama y un incremento del 10% en la incidencia en el año 2011. Sin embargo, en el análisis de una sola cohorte con seguimiento de por vida se estimó que el 4% de los cánceres detectados por cribado era sobrediagnosticado. La supervivencia de las mujeres detectadas por cribado aumentó en 2.5 años de media.

Todas las mamografías de cribado realizadas tuvieron un coste de 55.3 millones de Euros. El coste adicional de otros tests utilizados para el diagnostico en mujeres con mamografía positiva resultó en 12 millones de Euros. Aunque el diagnóstico precoz permitió ahorrar 39 millones de Euros en tratamientos, en términos de costes totales el escenario sin cribado tuvo un coste medio menor. En todas las simulaciones realizadas la ratio coste-efectividad incremental fue menor de 10,000€ por año de vida ajustado por calidad.

Los modelos de Gail y Chen mostraron buena calibración. El modelo de Barlow, sin embargo, sobreestimó el número de casos con cáncer. Aunque resultaron aceptables en

el test de calibración, los modelos de Gail y Chen sobreestimaron el número de casos para algunas categorías de densidad mamaria. Los modelos de Chen y Barlow con la proyección a 5 años fueron los que obtuvieron una mejor discriminación aunque no lo suficiente para poder aplicarlos en los programas de cribado.

CONCLUSIONES: Quince años después del inicio del programa de detección precoz del cáncer de mama, este estudio muestra un importante descenso en la mortalidad por cáncer de mama con daños limitados. Los efectos epidemiológicos del programa centralizado fueron confirmados por los resultados económicos que respaldan la continuación del programa de cribado de cáncer de mama en el País Vasco. No obstante, es necesario seguir optimizando el programa de cribado teniendo en cuenta el riesgo individual. En esta línea, los modelos de riesgo evaluados no resultaron herramientas útiles para la individualización de las estrategias de cribado.

1. Introducción

1.1. Epidemiología del cáncer de mama.

El cáncer de mama es el tumor maligno más frecuente entre mujeres de la Comunidad Autónoma del País Vasco (CAPV) (Tabla 1), región al noreste de España que limita con Francia. Concretamente uno de cada cuatro cánceres malignos diagnosticados en mujeres en nuestro territorio tiene su localización en la mama por lo que constituye un importante problema de salud pública (Izarzugaza et al. 2013).

Tabla 1: Localizaciones de cáncer más frecuentes en mujeres en la CAPV. Año 2009.

	LOCALIZACIÓN	%
1	Mama	26.5
2	Colon y recto	14.2
3	Cuerpo de útero	6.0
4	Sistema hematopoyético	5.7
5	Pulmón	5.5

La incidencia del cáncer de mama ha aumentado significativamente desde la introducción del cribado poblacional el año 1996 (Figura 1). Concretamente el número de casos diagnosticados ha pasado de 687 a 1226 en el periodo de 1986 al 2006. El incremento ha sido menos acusado en las tasas ajustadas por edad que han pasado de 65.0 a 86.9 en el mismo periodo (Izarzugaza et al. 2010). Más recientemente, en el año 2009, se diagnosticaron 1320 nuevos casos de cáncer de mama en mujeres de la CAPV, siendo la tasa ajustada de la incidencia de 88 casos por 100,000 mujeres al año (Izarzugaza et al. 2013).

La mortalidad desciende un 2.8% anual desde el año 1992 (Figura 1). La tasa ajustada de mortalidad (TAm) desciende de 26.0 en 1992 a 17.6 en el 2008, lo que indica mayor supervivencia relacionada con la mejora en los tratamientos y el avance en el diagnóstico (Izarzugaza et al.2010).

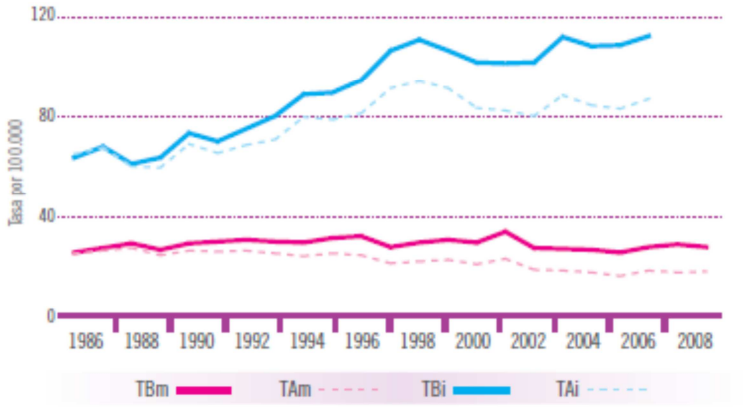


Figura 1: Evolución de la incidencia y mortalidad por cáncer de mama femenina por año en la CAPV.

A nivel de España se estima que en el año 2015 el cáncer de mama sea el cáncer con mayor incidencia en mujeres y el tercero en toda la población por detrás del cáncer colorrectal y el cáncer de pulmón (Figura 2). De hecho el 22% de los fallecimientos en mujeres en el año 2007 fue a causa de cánceres malignos y concretamente el 3.5% a causa del cáncer de mama (SEOM, 2010).

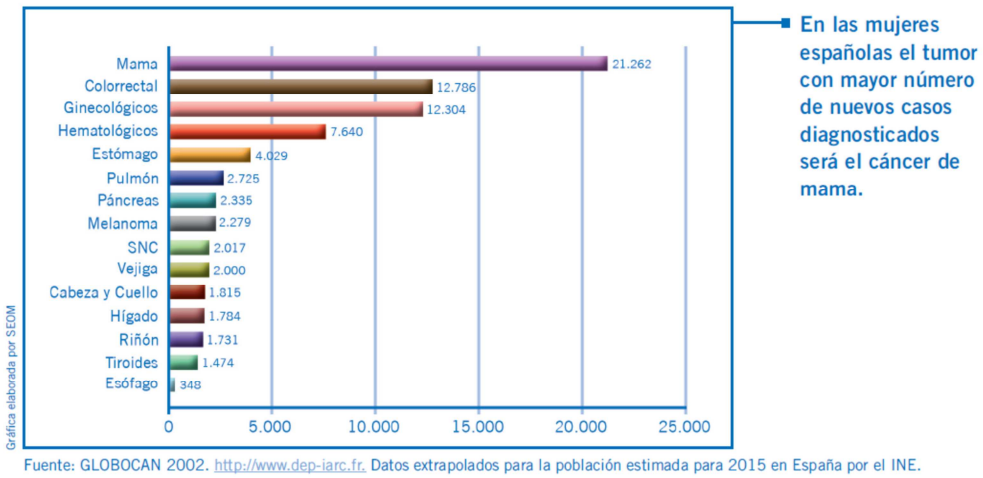


Figura 2: Incidencia de cáncer en mujeres en España por tipo de tumor. Estimado año 2015.

La clasificación de los cánceres de mama se hace en base al tamaño del tumor (T1-T4), a los nódulos (N0-N3) y ausencia o no de metástasis (M0-M1-MX), tal y como se describe en la tabla 2:

Tabla 2: Clasificación del cáncer de mama por tamaño, nódulos y metastasis.

TAMAÑO	T1	Tumor < 2 cm
	T2	Tumor entre 2 y 5 cm
	T3	Tumor > 5 cm
	T4	Tumor de cualquier tamaño que se extienda a pared torácica
NÓDULOS	**N0**	Ganglios axilares no palpables
	N1	Ganglios axilares móviles del lado del tumor
	N2	Ganglios axilares fijos en ausencia de metástasis
	N3	Metástasis a ganglios supra e infraclaviculares
METÁSTASIS	**M0**	Ausencia de metástasis
	M1	Metástasis a distancia
	MX	No se puede evaluar la metástasis

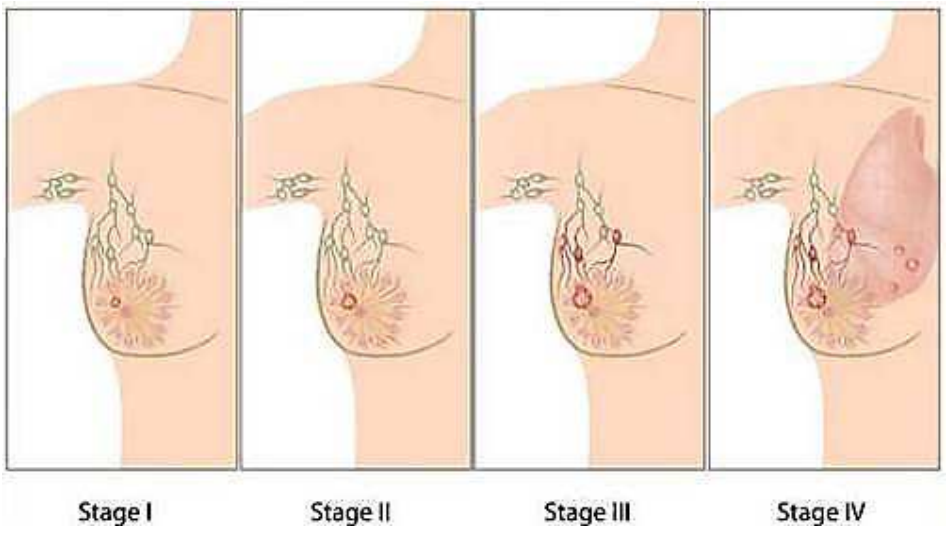


Figura 3: Estadificación del cáncer de mama

Tabla 3: Estadificación del cáncer de mama en relación a la clasificación por tamaño (T), nódulos (N) y metástasis (M).

Estadio I	T1 & N0 & M0	**Estadio III**	T1-2 & N2 & M0 T3 & N1 & M0
Estadio IIA	T1 & N1 & M0 T2 & N0 & M0		T4 & N0-1-2 & M0 T1-2-3-4 & N3 & M0
Estadio IIB	T2 & N2 & M0 T3 & N0 & M0	**Estadio IV**	T1-2-3-4 & N0-1-2-3 & M0

En los cánceres detectados en el País vasco previamente a que el programa de cribado poblacional se pusiera en marcha el 17% eran detectados en los estadios III y IV, los más avanzados.

1.2. Cribado de cáncer de mama.

Con estas cifras en mente, no cabe duda del importante problema de salud que supone actualmente el cáncer de mama en la población de mujeres. En estos momentos los factores de riesgo conocidos no son modificables por lo que no caben intervenciones poblacionales de prevención primaria. Hablamos de prevención primaria cuando el propósito de la intervención es limitar la incidencia de la enfermedad mediante el control de sus causas y de los factores de riesgo (Beaglehole et al., 1994). La prevención secundaria, en cambio no evita nuevos casos sino que persigue la curación de los pacientes diagnosticados y la reducción de las consecuencias más graves de la enfermedad mediante diagnóstico y tratamiento precoces. Los programas de detección precoz del cáncer de mama están incluidos entre las estrategias de prevención

secundaria. El cribado es un servicio de salud pública en el que los miembros de una población definida, que no necesariamente perciben tener un mayor riesgo, son invitados a someterse a pruebas diagnósticas para reducir el riesgo de la enfermedad o sus complicaciones (UK National Screening Committee, 2009).

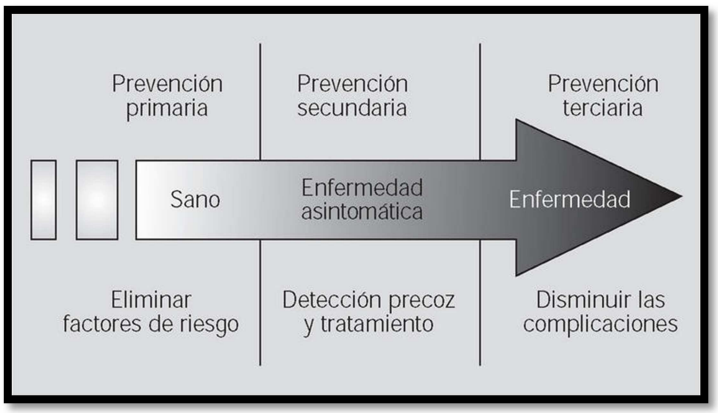


Figura 4: Prevención primaria, secundaria y terciaria

En el caso del programa para la detección precoz del cáncer de mama la intervención está generalmente dirigida a todas las mujeres de 50 a 69 años sin antecedentes familiares previos. Puesto que existen factores genéticos que aumentan el riesgo de desarrollar cáncer de mama, las mujeres con antecedentes familiares deberían tener un seguimiento distinto al de la población general puesto que tienen mayor riesgo.

Wilson y Jungner establecieron una serie de criterios necesarios para un cribado poblacional que aún se consideran referentes. El cribado de cáncer de mama cumple claramente todos ellos (Wilson and Jungner 1968):

1. Problema importante de salud.
2. Existencia de un tratamiento aceptado para los pacientes detectados.
3. Evidencia científica sobre el proceso diagnóstico y el tratamiento.
4. Periodo de latencia detectable.

5. Prueba adecuada, válida, fiable y eficiente.
6. Prueba aceptable para la población.
7. Enfermedad bien definida y con historia natural conocida.
8. Población diana bien definida.
9. El coste por caso detectado (incluyendo el diagnóstico y tratamiento de las personas detectadas) debe ser equilibrado.
10. Debe ser un proceso continuo y no una prueba puntual.

La prueba de cribado más utilizada es la mamografía, una exploración radiográfica de las mamas de la mujer realiza cada uno o dos años. Se trata de una prueba con gran validez diagnóstica, mínimos efectos adversos y bajo coste de aplicación por lo que posibilita llevar a cabo un cribado poblacional del cáncer de mama.

Como intervención de prevención secundaria, los programas poblacionales de detección precoz tienen como objetivo el adelanto del diagnóstico puesto que el pronóstico mejora considerablemente en los casos detectados en los estadios más leves. El cribado interrumpe la historia natural de la enfermedad y mejora el pronóstico con los tratamientos necesarios. Para tener resultados satisfactorios sin embargo es esencial, además, evitar el tratamiento de aquellos que no lo necesitan. Aunque la idea parezca sencilla la aplicación lleva a situaciones en las que es realmente difícil saber cómo actuar. Los pacientes en los que la detección precoz no suponga una mejora en su pronóstico sufrirán un periodo de morbilidad mayor por el adelanto diagnóstico. La detección de anomalías de pronóstico incierto o lesiones precursoras puede derivar en sobrediagnóstico y sobretratamiento (Márquez-Calderón, 2010).

La decisión de implantar el programa de cribado población fue tomada, en la mayoría de países desarrollados, en base a ensayos clínicos aleatorizados realizados entre 1963 y 1991.

En la revisión sistemática Cochrane llevada a cabo por Gotzsche (Gotzsche and Jorgensen, 2013) se concluye que tan sólo en tres de estos ocho ensayos se puede asegurar que la aleatorización de los grupos de intervención y control fue adecuada (Canada, Malmö y UK

age trial), además, se asegura que la aleatorización del The Edinburgh trial es totalmente inadecuada por lo que no se deberían tener en cuenta sus resultados.

Tabla 4: Ensayos clínicos aleatorizados que miden la efectividad del cribado de cáncer de mama mediante mamografía.

Año	Nombre del ensayo	Referencia
1963	New York trial / Health Insurance Plan trial	Shapiro, 1966
1976	Malmö trial	Andersson, 1981
1977	The Two-County trial	Tabar, 1985
1978	Malmö II trial	Andersson, 1997
1978	The Edinburgh trial	Anderson, 1986
1980	The Canadian trial	Baines, 1982
1981	The Stockholm trial	Frisell, 1986
1982	The Göteborg trial	Bjurstam, 1997
1991	The UK age trial	Moss, 1999

Cuando se analizan conjuntamente los resultados de estos ensayos clínicos, se muestra que en los ensayos aleatorizados de forma adecuada no se observa una diferencia estadísticamente significativa en cuanto a la mortalidad por cáncer de mama en el grupo de intervención y el de control con un riesgo relativo (RR) de 0.93 (IC 95% 0.79-1.09) a los 7 años de seguimiento y RR 0.90 (IC 95% 0.79-1.02) a los 13 años. Es más los resultados tampoco son estadísticamente significativos cuando se distinguen dos grupos según la edad (menos de 50 años y 50 años o más) de las mujeres intervenidas. Sin embargo, cuando el análisis tiene también en cuenta aquellos ensayos en los que la aleatorización es dudosa (7 ensayos en total), el riesgo relativo sí resulta estadísticamente significativo, RR 0.81 (IC 95% 0.72-0.90) a los 7 años y RR 0.81 (IC 95% 0.74-0.87) a los 13 años, aunque los resultados son menos fiables (Gotzsche and Jorgensen, 2013).

Los resultados positivos obtenidos en el New York trial (1963) y el Two County trial (1977) promovieron la implantación de los programas de cribado de cáncer de mama en Europa, sin embargo, este análisis muestra la incertidumbre a cerca de la fiabilidad de los resultados publicados puesto que en algunos casos se ha omitido información relevante a cerca de la metodología de aleatorización y en la mayoría se ha prestado poca atención a los efectos adversos de la intervención.

No sólo debemos tener en cuenta la incertidumbre mencionada en la interpretación de los resultados de los ensayos clínicos sino que también es cuestionable si estos resultados son aplicables a la situación actual. Es indudable que el enorme avance de los tratamientos de cáncer de mama en los últimos años ha aumentado significativamente la supervivencia de las mujeres afectadas, por tanto, la disminución en la mortalidad por cáncer de mama mostrado en algunos de los ensayos mencionados previamente podría verse diluido en el escenario actual donde los tratamientos aplicados son suficientes para la mejora del pronóstico de estos pacientes (Biller-Andorno, 2014).

Cabe destacar a su vez, que aunque en la mayoría de los casos el foco se centra en los resultados en cuanto a mortalidad por cáncer de mama o mortalidad general es igual de importante el análisis de los damos causados por el cribado poblacional de cáncer de mama. Cuando hablamos de daños o efectos negativos de la detección precoz hablamos sobre todo de casos sobrediagnosticados y mujeres que sufren ansiedad a causa de falsos positivos como resultado en las mamografías de cribado que implican también una serie de pruebas invasivas posteriores. La medida de casos sobrediagnosticados tiene grandes dificultades metodológicas dado que no es posible distinguir los tumores que son sobrediagnosticados y por tanto todos los cánceres detectados deben ser tratados de la misma forma. En realidad, se han llevado a cabo numerosos estudios con el objetivo de cuantificar su magnitud pero las cifras resultantes varían enormemente. En concreto en poblaciones en las que el cribado está extendido a toda la población objetivo y por tanto no existe un grupo control no es viable un estudio puramente observacional para la evaluación de este efecto adverso (Ascunce, 2015).

1.3. Programa de detección precoz de cáncer de mama en el País Vasco.

En los países más desarrollados el cribado de cáncer de mama es una de las principales intervenciones poblacionales ya que no cabe duda del beneficio que supone la detección precoz del cáncer a la hora del tratamiento y en la mejora de la supervivencia. Los primeros programas de cribado de cáncer de mama se iniciaron en los años 80 en Europa y a principios de los 90 en España. En 1995, el Departamento de Sanidad del Gobierno Vasco puso en marcha en el Área Sanitaria de Araba el Programa de Detección Precoz del Cáncer de Mama (PDPCM), extendiéndose a lo largo de 1997 a toda la CAPV (Sarriugarte, 2011). El test diagnostico del cribado consistió en la realización de una mamografía bilateral en doble proyección: craneo-caudal y oblicua medio-lateral con periodicidad bienal a las mujeres entre 50 y 64 anos y que se extendió hasta los 69 anos a partir de 2006 de forma progresiva (evitando dar bajas por edad a mujeres nacidas en 1941 y anos posteriores), completándose la ampliación hasta los 69 anos, en el ano 2010.

Figura 5: Instante de la realización de una mamografía.

Un panel de expertos convocado por el Servicio de Evaluación de Tecnologías Sanitarias del Departamento de Sanidad del Gobierno Vasco, aconsejo un modelo de sistema organizativo basado en una unidad de detección en la que se incluyese, tanto la exploración mamográfica básica, como la realización de un segundo estudio radiológico cuando este fuera preciso. Para la investigación citológica, el diagnóstico y el tratamiento de las lesiones detectadas recomendaron la derivación de los casos sospechosos a

Unidades Hospitalarias de Patología Mamaria (unidades diagnostico-terapéuticas). Por otra parte el Plan Integral de Prevención y Control del Cáncer en Euskadi 1994-1998, al recomendar la puesta en marcha de un Programa de Diagnostico Precoz del Cáncer de Mama (PDPCM) aconsejo el aprovechamiento de los recursos existentes en la red sanitaria publica de la CAPV y por tanto la integración de los recursos necesarios para la ejecución del programa en la red asistencial.

La organización funcional del PDPCM se estableció en tres niveles: el nivel de gestión, el de detección y valoración de lesiones sospechosas de cáncer de mama y por ultimo el nivel diagnostico-terapéutico.

1. Unidad de gestión: Esta formada por una unidad central, cuyas funciones son:
   - Realizar la planificación en función de los objetivos marcados en el PDPCM.
   - Implantación del PDPCM.
   - Gestionar los flujos de información, recursos humanos y materiales.
   - Evaluar el desarrollo de las actividades.
   - Supervisar el cumplimiento de los objetivos.
   - Asegurar el cumplimiento de los niveles de calidad.
   - Coordinar el PDPCM: estableciendo los criterios de actuación en aspectos organizativos, velando por la uniformidad de criterios en todo el ámbito y niveles del PDPCM, garantizando la coordinación e integración de los diferentes niveles tanto de gestión como asistenciales.

2. Unidades de detección-valoración compuestas a su vez por dos escalones
   a. Unidad de detección: unidad de radiología para la realización de mamografías y ecografías, cuyas funciones son:
      - Detección de lesiones sospechosas de cáncer de mama en un primer nivel de estudio mediante la realización de mamografía, y en los casos en los que se requiera un segundo estudio, la realización de nuevas mamografías (nuevas proyecciones, magnificaciones,...) y/o la realización de ecografías, mantenimiento y control de calidad técnica de la unidad.

b. Unidad de valoración: unidad dotada de radiólogo entrenado en técnicas de detección precoz de cáncer de mama, para la lectura diagnostica de las mamografías realizadas en las unidades de detección asignadas a su unidad, cuyas funciones son:

• Lectura diagnostica de las mamografías procedentes de la unidad de detección, realización de un segundo estudio (nuevas proyecciones, magnificaciones y/o ecografía en aquellos casos que lo requieran), realización de informes, canalizar la información a la unidad central, derivación de los hallazgos sospechosos a la unidad diagnostico-terapéutica, participación en la toma de decisiones sobre el seguimiento de los casos derivados a la unidad diagnostico-terapéutica mediante reuniones periódicas con los miembros de la unidad de patología mamaria de su hospital de referencia.

3. El tercer nivel lo constituyen las unidades diagnostico-terapéuticas ubicadas en los hospitales de referencia y cuyas funciones son:

• Diagnostico de las lesiones sospechosas derivadas desde la unidad de detección-valoración.

• Tratamiento de la patología maligna detectada.

• Seguimiento de los casos que han precisado tratamiento.

• Comunicación de los casos detectados a la comisión de tumores del propio hospital, que los derivara al Registro Hospitalario de Cáncer.

Tal como se ha señalado y a diferencia de los programas de otras comunidades autónomas, en la CAPV la gestión del cribado se descentralizo en diferentes hospitales y ambulatorios y no se configuró como un centro de gasto. La consecuencia de esta decisión fue que el programa no ha gestionado directamente los recursos utilizados en la realización del cribado. La red de atención especializada de Osakidetza de Bizkaia y Araba y Onkologikoa en Gipuzkoa asumieron la ejecución del mismo. Desde el punto de vista de la evaluación del programa, la falta de una estructura propia de contabilidad ha dificultado el conocimiento sistemático de los costes del servicio prestado por el programa (Finkler, 1982). En el ano 2003 se llevo a cabo un análisis de costes del

programa dentro del informe que evaluó la ampliación de la edad hasta los 69 anos (Gutierrez, 2004).

Los recursos actuales asignados al PDPCM para la fase de cribado, lo componen 10 unidades de detección, todas ellas unidades fijas (no hay unidades móviles en el programa), de las que ocho son dependientes de Osakidetza y dos son concertadas con Onkologikoa, y seis unidades de valoración (cinco de Osakidetza y una de Onkologikoa).

Las descripciones de las actividades del PDPCM se recogen en el denominado proceso principal del programa. Este proceso comienza con la actividad de citación y concluye con la recepción del informe por parte de la mujer participante en el PDPCM. Este proceso a su vez puede ser dividido en tres subprocesos que se refieren a la actividad de citación, la detección de lesiones sospechosas y en tercer lugar el seguimiento de las lesiones detectadas y la emisión de informes correspondientes (Sarriugarte, 2011).

Actividad de citación: Básicamente consiste en la obtención de datos sobre las mujeres que constituyen la población objetivo del programa a partir de los listados de los padrones municipales que cruzados con datos de diagnostico previo de cáncer de mama y de mortalidad, permiten obtener el listado de la población diana. El acceso a las mujeres invitadas al programa se lleva a cabo mediante el padrón ya que todas las mujeres empadronadas mayores de 49 anos entran en el PDPCM de la CAPV, siendo el único motivo de exclusión, el tener diagnosticado un cáncer de mama. Tras generar un calendario de citas, que previamente ha sido consensuado con la citada unidad, se emite la carta-cita que se envía al domicilio de la mujer. En dicha carta consta el día, hora y lugar al que debe acudir, así como el teléfono de contacto para aclarar cualquier duda o realizar algún cambio.

Detección de lesiones sospechosas: Las unidades de detección-valoración llevan a cabo la detección de las lesiones sospechosas mediante la realización de mamografías bilaterales en dos proyecciones (cráneo-caudal y oblicua-medio-lateral). En algunos casos es necesaria una revaloración (otras proyecciones, magnificaciones, compresión localizada, ecografías) para la correcta valoración por parte del radiólogo. En los casos de alteraciones cuya agresividad deba ser descartada mediante técnicas complementarias o

en los casos en que exista una imagen altamente sospechosa de malignidad, se les cita en el hospital de referencia, donde proceden al diagnostico y el tratamiento de las lesiones detectadas.

Emisión de informes: A las mujeres que no requieren control y se remiten a la siguiente vuelta del programa, se les envía una carta informe recomendándoles volver a realizarse una nueva exploración radiológica a los dos años, mientras que a las que requieren un control radiológico antes de los dos años, se les envía una carta informe recomendándoles un control antes de los dos años (seis o 12 meses).

Concluido el protocolo terapéutico, en los casos en que tras las pruebas diagnosticas el resultado es de benignidad y, se toma la actitud de vuelta al programa, se envía a la mujer además del informe de alta del hospital, una carta, especificando la fecha recomendada para realizar el siguiente control.

En el caso de diagnostico de cáncer, se continua con la pauta terapéutica y se realiza el seguimiento de la paciente en el hospital de referencia, comunicándose el diagnostico al registro de cáncer. La mujer es dada de baja del programa para siguientes vueltas, aunque el programa recoge los datos generados en la actividad diagnostico-terapéutica y de seguimiento para su posterior evaluación.

1.4. Evaluación de intervenciones de salud pública.

El alto coste de las intervenciones de salud pública y la complejidad de su gestión hacen indispensable la evalución en términos de recursos económicos, materiales y humanos. El presupuesto no es infinito y la demanda poblacional aumenta conforme crece la oferta en materia tecnológica (Garcia-Altes, 2011). En el ámbito de la salud, la evaluación económica es la herramienta que permite comparar diferentes alternativas en términos de coste y efectividad (Drummond, 2005; Gold, 1996). El objetivo final es dar un servicio de calidad al conjunto de la población, sin embargo, es importante optimizar los recursos disponibles y por tanto sacar la mayor rentabilidad de los mismos. En este

contexto es fundamental la definición de un marco estándar para la evaluación de intervenciones de salud pública que permita identificar el programa que más beneficio aporta a la población en relación a la inversión necesaria.

Se considera beneficio todo aquello que contribuye a aumentar el bienestar de las personas. En el caso de un tratamiento médico o de un programa sanitario se pueden agrupar en dos grandes grupos (Pinto, 2011):

1. Beneficios sanitarios: Incluye las mejoras en la salud causadas por una intervención.
2. Beneficios no sanitarios:
    a. Mejoras en la productividad: como efecto indirecto de la mejora en salud el paciente disminuye el período de convalecencia y aumenta el tiempo dedicado al trabajo. Puede ser incluso que no aumente el tiempo de trabajo pero, sin embargo, este tiempo resulte más efectivo debido a la mejora de las condiciones físicas o mentales.
    b. Mejoras en la calidad de vida: beneficios que no se plasman en un efecto directo en términos de salud ni ganancias monetarias pero sin embargo mejoran el día a día de la persona.

Todos estos conceptos de beneficio de las decisiones clínicas son características valoradas desde el punto de vista de los pacientes por lo que el beneficio está directamente relacionado con las preferencias individuales. En la evaluación de programas de salud pública los beneficios y costes deben adoptar una perspectiva social, no individual.

Una manera de medir el beneficio de un programa sanitario sin recurrir a medidas naturales ni unidades monetarias es mediante la calidad de vida relacionada con la salud (CVRS). La calidad de vida es un concepto muy amplio por lo que se utiliza el concepto CVRS cuando hacemos referencia a aquellos aspectos de la calidad de vida que están directamente relacionados con la salud. Estas medidas deben tener ciertas propiedades que permitan la comparación de distintas intervenciones y su interpretación en relación a la esperanza de vida. Conocidos la cantidad de años de vida (Y) y la calidad de vida (Q) el

estado de salud de un individuo está completamente definido por el par (Q, Y) mediante los años de vida ajustados por calidad (QALY). Normalmente una salud perfecta está asociada al valor Q=1 por lo que un QALY se interpreta como un año de vida en un estado de salud perfecto. La utilidad establecida para el estado de la muerte es Q=0 por lo que cuando la calidad de vida toma valores negativos se asume que el individuo se encuentra en un estado de salud peor que la muerte.

Tabla 5: Tipos de estudios de evaluación económica

	Minimización de costes	Coste-efectividad	Coste-utilidad	Coste-beneficio
Medida de los costes	Unidades monetarias	Unidades monetarias	Unidades monetarias	Unidades monetarias
Efectividad	Idéntica	Común a las alternativas	Supervivencia y CVRS	No común a las alternativas consideradas
Medida de resultados	No procede	Unidades naturales de las alternativas	AVACs	Unidades monetarias
Estrategia de análisis	Comparar el coste de las alternativas	Comparar el coste por unidad de resultado de las alternativas	Comparar el coste por AVAC en las alternativas	Comparar las razones coste-beneficio de las alternativas
Criterio de elección	Alternativa de menor coste	Alternativa con menor coste por unidad de resultado ganado	Alternativa con menor coste por AVAC ganado	Alternativa con menor cociente coste-beneficio o mayor beneficio neto

En las intervenciones poblacionales en las que la perceptiva social toma un valor añadido es necesario tenerlo en cuenta también en la selección de las preferencias. Lo más adecuado sería conocer la valoración de la población general con respecto a los estados de salud involucrados en la evaluación. En general, al optar por medidas indirectas de utilidades de los estados de salud a partir de los sistemas de clasificación multiatributo, como el cuestionario Euroqol 5D (EQ-5D), se está siguiendo esta pauta dado que estos sistemas obtienen puntuaciones relativas a los diferentes atributos y sus niveles a partir de preferencias de la población general.

En la evaluación de intervenciones de salud pública se pueden remarcar cuatro puntos fundamentales (Weatherly et al., 2009):

1. El efecto de la intervención:

A la hora de evaluar de tecnologías sanitarias dirigidas a grupos de individuos con características específicas está claro que los ensayos clínicos aleatorios son los estudios de referencia para la comparación de los efectos de diferentes alternativas. En las intervenciones de salud pública, sin embargo, es difícil llevar a cabo ensayos clínicos aleatorizados dado que la población objetivo es generalmente toda la población. En estos casos deberíamos aplicar nuevas metodologías que permitan obtener estimaciones insesgadas de los efectos de la intervención. Esta es una tarea complicada puesto que es muy costoso llevar a cabo estudios que reflejen los resultados a largo plazo que se espera obtener en las intervenciones de salud pública. Incluso cuando se dispone de ensayos clínicos aleatorizados es innegable el problema que supone el análisis de la proyección a largo plazo de los efectos de la intervención.

Está claro que en la evaluación de intervenciones de salud pública es necesario tener en cuenta toda la evidencia disponible. Los ensayos clínicos podrían demostrar el efecto a corto plazo con medidas intermedias que pudieran estar relacionados con los efectos a largo plazo en estudios observacionales de mayor longitud. De hecho, la evaluación de intervenciones farmacológicas y quirúrgicas mediante ensayos clínicos es una práctica habitual en algunos estados a nivel de Europa. Sin embargo, en las intervenciones de salud pública los ensayos clínicos no son habituales dadas las limitaciones ética, económicas y de tiempo, en consecuencia, su evaluación tampoco es frecuente. Además de los ensayos clínicos, son necesarias también las técnicas estadísticas avanzadas como el propensity score, análisis de series temporales en experimentos naturales, modelos econométricos más sofisticados o incluso modelos matemáticos de simulación que reproduzcan la historia natural, dado que su aplicación ayuda en búsqueda de la evidencia. De hecho, estos modelos matemáticos pueden servir para sintetizar los resultados observados en diferentes estudios donde la población a estudio es diferente a la población local y permiten proyectar a costes y beneficios a largo plazo los resultados intermedios observados (Briggs et al., 2006).

2. Medida y valor de los resultados:

Para la evaluación de los resultados a largo plazo de una intervención es muy importante la proyección a largo plazo mencionada anteriormente. Habitualmente los resultados se miden en años de vida ajustados por calidad lo cual además de la proyección de los resultados exige la clasificación de los estados de salud futuros. En la evaluación de intervenciones de salud pública los efectos se deben considerar de una manera más amplia ya que puede haber efectos que no estén relacionados directamente con la salud y por tanto sean difíciles de convertir en QALYs como puede ser, por ejemplo, un mayor nivel educativo. Es más, los efectos de intervenciones poblacionales pueden beneficiar también a parte de la población que no ha sido directamente intervenida.

Es importante tener en cuenta tanto los beneficios como los daños ocasionados por la intervención así como los efectos que no están relacionados con la salud. La conversión de dichos daños y beneficios a QALYs limita el rango de resultados en salud que se tienen en cuenta. Indudablemente la planificación de un programa poblacional implica que la toma de decisiones se debe hacer en base a múltiples criterios con el objetivo de maximizar los beneficios al mismo tiempo que se asegura que los daños causados serán mínimos. De hecho, la toma de decisiones multi-criterio se hace en base a las preferencias del decisor quien subjetivamente asigna el peso que considera a cada uno de los resultados (positivos o negativos) de la intervención (Baltussen et al., 2006).

3. Costes y consecuencias intersectoriales

El impacto de los programas de salud pública puede implicar un amplio rango de áreas. Tanto los costes como los efectos de la intervención pueden estar asociados a diferentes áreas de sector público no únicamente el área de salud. Puede ser, por ejemplo, que el aumento del presupuesto del departamento de salud suponga ahorros para el departamento de servicios sociales debido a la reducción de la población en estado dependiente o puede ocurrir también a la inversa, que la inversión en centros gerontológicos en los que mejora el cuidado a los ancianos residentes en la misma reduzca el consumo de recursos en sanidad pública de estas personas. Es necesario en el marco de evaluación de dichas intervenciones considerar estos aspectos intersectoriales

en caso de ser relevantes. Sin embargo, en la practica no existe un marco intersectorial en el que se utilice una medida genérica común para la medición de los resultados en diferentes ámbitos, ni siquiera la posibilidad de transferencias compensatorias entre diferentes departamentos lo que dificulta la creación de un marco teórico general.

4. La equidad

En cuarto lugar y último, pero no menos importante, está la equidad. De hecho es fundamental para que una intervención de salud pública cumpla su propósito de mejorar la salud a nivel poblacional que se cumpla el criterio de equidad. La salud pública es el área encargada no sólo de la producción y distribución de los cuidados en salud si no que también de la producción y la distribución de la salud (Maynard, 2012). Dicho de otra forma, debe garantizar la equidad tanto en la distribución del acceso a servicios sanitarios como en el beneficio que las intervenciones de salud pública producen en la población para los diferentes niveles socioeconómicos.

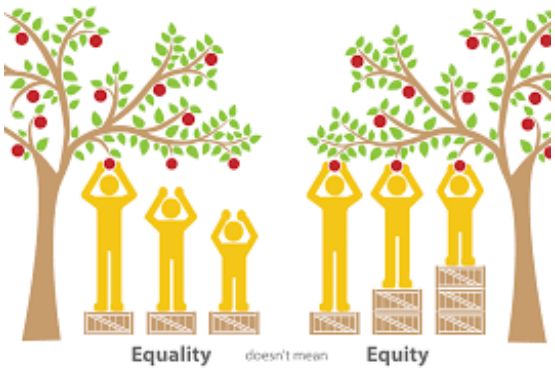


Figura 6: Representación gráfica de la equidad.

Este es un aspecto claramente diferenciador de las intervenciones de salud pública que pocas veces es referenciada en la evaluación de los programas poblacionales. De hecho, como ya mencionábamos en el segundo punto es difícil interpretar en términos de QALYs el beneficio de la equidad, dado que es un resultado que no está directamente relacionado con la salud, sino con beneficios sociales. Una de las principales asunciones

cuando se aplican los QALYs como medida de beneficio es que el valor de un QALY no depende de quién reciba dicho beneficio. Realmente, puede que el único objetivo o el objetivo principal de una intervención sea precisamente que la distribución de los servicios de salud sea más equitativa.

Al mismo tiempo es importante saber el peso que los decisores dan a la equidad en el contexto de las intervenciones de salud pública, es decir, cual es el beneficio poblacional que están dispuestos a perder en beneficio de una distribución equitativa de los resultados del programa a implantar.

La dificultad de medir los efectos de intervenciones poblacionales lleva a la implantación de programas con un gran coste en base a escasa evidencia. De hecho, la evidencia de la efectividad de la intervención es necesaria pero no suficiente para su implantación a nivel poblacional dado que puede ser efectiva pero no ser coste-efectiva (Maynard, 1997). Toda decisión tomada sobre este tipo de intervenciones implica un coste de oportunidad, esto es la limitación de recursos supone que la cantidad invertida en una opción no estará disponible para llevar a cabo otra de las opciones y como consecuencia la población no podrá beneficiarse de los efectos de las intervenciones no financiadas (Maynard, 2012).

1.5. Evaluación económica

Los recursos son escasos y aunque cada vez el presupuesto destinado a Salud es mayor, la necesidad tiene que ser limitada, es decir, es necesario decidir optimizar los beneficios del dinero invertido (Sacristán, 2004). La evaluación económica de intervenciones sanitarias se define como la comparación de dos opciones alternativa en términos de costes y consecuencias (Drummond, 2005). Se consideran opciones alternativas los diferentes modos de abordar el problema u organizar los recursos disponibles con el objetivo de mejorar la salud de los pacientes.

En realidad, el término Evaluación Económica engloba varias técnicas o procedimientos que pueden emplearse para comparar información sobre la relación que existe entre el coste y los resultados de las intervenciones destinadas a la mejora de la salud de los individuos (Prieto, 2004). Las técnicas de evaluación económica utilizan la teoría económica como base en la priorización de las alternativas analizadas. El criterio que se aplica es el de la eficiencia, entendiendo por eficiente el proceso que más salud produzca con los recursos dados o aquel que menos coste tenga entre los que producen el mismo resultado (Drummond, 2005).

Son tres las principales características que definen la evaluación económica: la medida de la efectividad, la medida de los costes y el comparador utilizado para la elección final. Los aspectos relacionados con la medida de la efectividad ya han sido mencionados en el apartado anterior por lo que en esta sección nos centraremos en el cálculo de los costes y los criterios utilizados para la elección final.

1.5.1. Costes sanitarios

El análisis comparativo de los costes asociados a las intervenciones alternativas es común a todas las formas de evaluación económica. El coste de un recurso está definido por la cantidad total de recurso consumido y el valor monetario de la utilización de una unidad de ese recurso (Drummond, 2005). En el cálculo de los costes atribuible a una intervención es primordial identificar los recursos más relevantes de las opciones que se están comparando.

Generalmente, los costes se clasifican teniendo en cuenta si son costes directos o indirectos por una parte y si se trata de costes del ámbito sanitario o no sanitario (Johnston, 1999). Los costes directos son esencialmente transacciones monetarias que implican servicios y productos tanto sanitarios como no sanitarios. Los costes sanitarios que por lo común se consideran están directamente relacionados con la intervención concreta evaluada e incluyen los costes de hospitalización, tratamiento, honorarios

profesionales, pruebas de laboratorio, rehabilitación y equipo médico, entre otros. Los gastos que son consecuencia directa de la enfermedad pero no implican la compra de servicios sanitarios, se consideran como costes directos no-sanitarios, puesto que suponen un impacto importante para las finanzas del paciente y sus familiares. Finalmente los costes indirectos incluyen los costes producidos por la morbilidad o la mortalidad prematura asociada a la enfermedad. Desde el punto de vista social deberían ser considerados en la evaluación económica de cualquier intervención a nivel poblacional aunque no es habitual incorporarlo.

Tabla 6: Clasificación general de costes

	Sanitarios	No sanitarios
Directos	Cuidados hospitalarios, tratamientos farmacológicos,…	Gastos de desplazamiento, cuidados informales,…
Indirectos	Consumo de servicios sanitarios a lo largo de los años de vida ganados a causa de la intervención	Pérdida de productividad, Coste de oportunidad del tiempo invertido en el tratamiento,…

La perspectiva del análisis define generalmente la elección de los costes más relevantes para el estudio. La perspectiva más extendida es la de la sociedad que implica que toda inversión de recursos que suponga un coste de oportunidad para cualquier elemento de la sociedad debe tenerse en cuenta. Desde la perspectiva del sistema sanitario los costes serán restringidos a aquellos que correspondan a los servicios de salud prestados por el sistema en cuestión. Según las recomendaciones del National Institute for Clinical Excellence (NICE) la evaluación económica de intervenciones sanitarias se debería llevar a cabo desde la perspectiva del Sistema Nacional de Salud (NICE, 2001).

Una vez identificados los costes, se debe medir o estimar la cantidad de recursos necesaria primero para posteriormente calcular su valoración en unidades monetaria. En

el primer paso, cada elemento de coste debe ser expresado en unidades naturales, tales como, número de pruebas diagnósticas, días de hospitalización, tiempo del paciente o de los familiares, etc. Los costes serán calculados como el producto de un vector de cantidades de recursos y un vector de precios unitarios de tales recursos (Pinto, 2011). La medida de los costes tiene que ver con el vector de cantidades, mientras que la valoración se refiere a la asignación de precios unitarios a tales cantidades.

La medida de los recursos utilizados a consecuencia del uso de la tecnología que se quiere evaluar puede llevarse a cabo de forma paralela a la recogida de datos de efectividad en un ensayo clínico o en su defecto utilizar información retrospectiva para su estimación. Si el estudio de costes se realiza dentro del ensayo clínico aleatorizado se podrá disponer de datos individualizados sobre el uso de recursos en cada paciente y esto permitirá a su vez relacionar el consumo a las características de los individuos a estudio. Sin embargo, los ensayos clínicos generalmente no reflejan la práctica clínica habitual, por lo que la estimación de los costes puede ser diferente a los costes que se producirían en un escenario real. La alternativa es la utilización de información retrospectiva junto con la opinión de expertos para la estimación del coste, eso sí, en estos caso el coste no puede ser asociado a las características del paciente, únicamente se determina el coste medio (Pinto, 2011). Además de estos procedimientos, herramientas como los Grupos Relacionados por Diagnóstico (GRD) permiten también valorar directamente el coste de la intervención obviando las fases de medición y valoración. Los GRD proporcionan estimaciones de costes medios basadas, en principio, en la medición y valoración de costes sobre grupos numerosos de pacientes.

1.5.2. Análisis coste-efectividad

El análisis coste-efectividad es el método mas utilizado entre los diferentes métodos que se distinguen en la evaluación económica. En este tipo de análisis se tiene en cuenta el coste total de la intervención, incluyendo todos los recursos utilizados en el proceso,

por una parte, pero también su efecto en la calidad de vida de los pacientes intervenidos (Briggs, 2006). Es más, determina de forma numérica cual es la relación entre los costes de una intervención dada y las consecuencias de esta. El valor relativo de la intervención se expresa habitualmente como el cociente que se obtiene al dividir el coste neto de la intervención por su beneficio neto o efectividad (Drummond, 2005).

En el hipotético caso de que no hubiera limitación de recursos todas las intervenciones que produjeran beneficios deberían ser aplicados, sin tener en cuenta su coste. Sin embargo, todos los sistemas sanitarios tienen un presupuesto limitado que obliga a priorizar en la elección de intervenciones beneficiosas. El objetivo en la gestión del presupuesto será obtener el mayor beneficio poblacional posible con los recursos disponibles. El análisis coste-efectividad permite comparar diferentes intervenciones y hacer una clasificación ordenada de ellas en función de la relación existente entre su coste y su efectividad (Prieto, 2004).

Principalmente el indicador que se utiliza en la comparación de las intervenciones dos a dos es la ratio coste-efectividad incremental (ICER). Este valor compara los costes y efectos de una intervención con otra intervención que puede estar dirigido a la misma enfermedad o no, pero ambos deben medir sus resultados en las mismas unidades (Drummond, 2005). Generalmente, la intervención a estudio es comparada con la intervención más utilizada en la práctica clínica hasta ese momento, ambos con el mismo objetivo terapéutico. En el caso de dos intervenciones alternativas, A y B, el ICER se calcula de la siguiente forma:

$$ICER = \frac{C_A - C_B}{E_A - E_B}$$

Donde C_A y C_B son los costes de las alternativas A y B y las medidas de efectividad correspondientes se muestran como E_A y E_B. Por tanto el ICER informa del coste adicional por unidad de beneficio también adicional con respecto a la alternativa basal. En la interpretación del ICER, las intervenciones con el valor más bajo tienen un menor coste por cada unidad de beneficio neto o efectividad que producen, los de valor más alto en cambio, son menos eficientes con respecto a la alternativa basal seleccionada.

Es imprescindible además, en la priorización de las intervenciones, tener en cuenta no sólo la relación entre el coste y los resultados de la intervención sino que también las preferencias temporales de la población. En economía se conoce como preferencia temporal la desigual valoración que hacen los sujetos de las ganancias y de los costes en función de cual sea el momento en el tiempo en el que se producen (Pinto, 2011). Esa conducta temporal está demostrado que se cumple tanto en el caso de los beneficios como de los costes, valorando menos los incurridos en el futuro. Es importante aclarar que la preferencia temporal no está relacionada con la inflación y sería necesario aplicarlo incluso en el caso de tener precios constantes en el tiempo e intereses bancarios nulos. Más concretamente es debido a la incertidumbre relacionada a eventos futuros, al desconocimiento sobre si estos llegarán a ocurrir, por lo que las personas tienden a valorar más los resultados instantáneos.

Esta dimensión temporal de las preferencias se debe incorporar a la evaluación económica. Cuando dos programas tienen sus costes y beneficios repartidos a lo largo del tiempo es necesario introducir ajustes basados en la preferencia temporal, de modo que sean comparables. Para homogeneizar las magnitudes que componen un flujo de costes o beneficios que se extiende a lo largo de sucesivos períodos se recurre al procedimiento del descuento, que consiste en expresar todos los valores futuros en su valor equivalente en el momento actual.

Para obtener el valor actual de un ingreso o un pago futuro, debemos multiplicar el valor futuro por la expresión $\frac{1}{(1+r)^t}$ donde t es el número de ciclos (generalmente años) transcurridos desde el momento presente hasta el momento en el que ocurre el ingreso o el pago futuro y r es la tasa de descuento. A la expresión anterior se la denomina factor de descuento. Este factor de descuento es adecuado cuando el coste futuro es un coste puntual, pero cuando el coste se prolonga a lo largo de un intervalo (t_1, t_2) el factor de descuento se calcula con la siguiente expresión:

$$\int_{t_1}^{t_2} \exp(-rx)\, dx = \frac{1}{r}(\exp(-rt_1) - \exp(-rt_2))$$

Se puede verificar fácilmente que cuando el intervalo es $(0, t)$ la expresión del factor de descuento continuo toma valores muy similares a la expresión utilizada en el caso del tiempo discreto o ciclos.

$$\frac{1}{(1+r)^t} \sim \frac{1}{r}(1 - \exp(-rt))$$

Así como existe unanimidad en relación a la necesidad de descontar o actualizar los costes futuros, el descuento de los beneficios futuros en la evaluación económica de intervenciones sanitarias es aún un punto de discusión. En este aspecto, el NICE recomienda aplicar la misma tasa de descuento del 3% tanto a los costes como a los resultados de efectividad en el análisis coste-efectividad (NICE, 2001).

1.5.3. Análisis del impacto presupuestario.

Los estudios coste-efectividad permiten cuantificar el beneficio añadido que supone un nuevo tratamiento en relación con su coste y compararlo con lo que representan otros medicamentos ya aceptados. En este sentido, proporciona un marco teórico cuyo éxito ha sido reconocido internacionalmente. El crecimiento de su peso en la literatura científica así lo atestigua. Sin embargo, algunos autores han señalado que su aplicación en el proceso de toma de decisiones de la vida real plantea muchas dificultades y que su uso en la práctica ha sido menor del que cabría esperar por su éxito como método científico. Una de las quejas que se han expresado es la dificultad de trasladar sus resultados al contexto de un sistema de gestión sanitaria centrado en el presupuesto. En respuesta a estas críticas, desde hace años se utiliza una herramienta complementaria denominada análisis del impacto presupuestario (AIP) dirigida a estimar la repercusión de la introducción de un nuevo medicamento en los presupuestos de los años venideros.

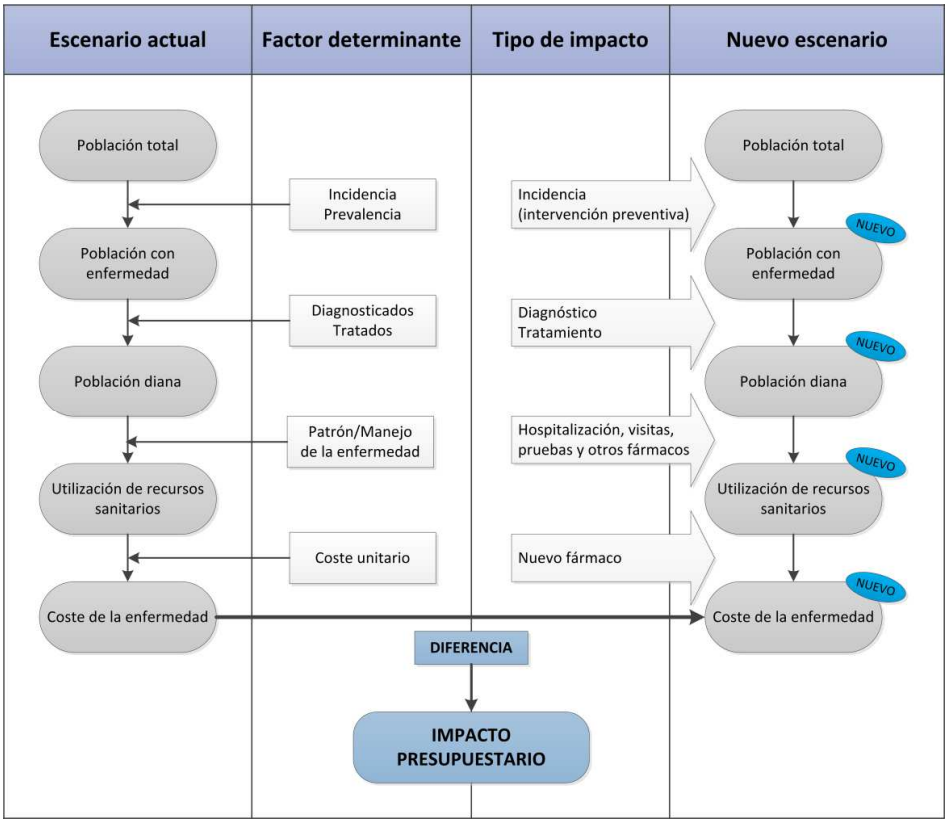


Figura 7: Efecto de una intervención sanitaria en el presupuesto

Según Mauskopv (Mauskopv, 2007), el AIP mide el impacto de un nuevo tratamiento en el coste anual, el beneficio en salud anual y en otros resultados de interés en los años posteriores a su introducción en un sistema nacional de salud o en un plan de salud privado. A pesar de ser usado de forma casi sistemática en los procesos de registro de medicamentos, su presencia en la literatura científica ha sido testimonial. La causa principal es que el formato de AIP utilizado ha consistido en general en modelos sencillos basados en supuestos provenientes de la literatura y con frecuencia en opiniones de expertos. En los últimos años diferentes autores han propuesto guías para su desarrollo con requisitos más exigentes y que en la medida que se generalicen van a dotar al AIP de rango científico. Otra limitación importante de este tipo de estudios ha sido que la perspectiva desde la que se han llevado a cabo era con frecuencia la del impacto en

farmacia exclusivamente. Las perspectivas que se aceptan como estándares en las guías de evaluación económica exigen tener en cuenta la repercusión de la introducción del nuevo tratamiento en el consumo de recursos en el sistema sanitario en su conjunto o en la sociedad en general. Sin embargo, las dificultades metodológicas para representar los cambios de la historia natural de una enfermedad a nivel poblacional han empujado a que la perspectiva se ciñese a la de farmacia y a que su uso se limitase al puramente administrativo.

El AIP es distinto según la perspectiva que se utilice. En función de la perspectiva seleccionada para la evaluación económica el AIP incluirá únicamente algunos o todos los tipos de coste relacionados con la intervención a estudio (Pinto and Sanchez-Martinez, 2010). Al analizar las ventajas y desventajas de los diferentes enfoques aparece claramente la contradicción entre el marco teórico de la evaluación económica y el uso habitual que se hace de los AIP. Todas las recomendaciones de los paneles de expertos señalan la perspectiva del conjunto de la sociedad como la más adecuada. Sin embargo, la mayoría de los estudios coste-efectividad utilizan la perspectiva del sistema sanitario y en el caso de los AIP el enfoque más frecuente es el de farmacia.

Uno de los problemas que dificultan la adopción de una perspectiva global es la dificultad que plantean los métodos requeridos para la estimación de los resultados finales de la intervención. La fuente que se dispone habitualmente son los datos de eficacia de los ensayos clínicos. Con frecuencia, éstos son resultados intermedios que no permiten predecir el impacto a medio plazo de la nueva intervención en estados de salud con necesidades de cuidados específicas. Los modelos de Markov que son de uso rutinario en los estudios coste-efectividad están dirigidos al seguimiento de una cohorte de pacientes para calcular su esperanza de vida en cada estado, cuantificar el coste incremental y la efectividad incremental y poder obtener así el coste por AVAC ganado con el nuevo tratamiento. La obtención de la prevalencia poblacional obliga al uso de modelos como la simulación con eventos discretos cuyo uso está poco extendido en la literatura científica. En este sentido la generalización de este tipo de modelos va a proporcionar estimaciones más acordes con la realidad social del impacto presupuestario de los nuevos tratamientos. La falta de modelos matemáticos que sean capaces de cuantificar en

términos de coste los cambios en el consumo de recursos sanitarios a medio plazo hace que los análisis del impacto que representa el tratamiento de la hepatitis C se limiten a la perspectiva de farmacia. En este sentido, la falta de medición facilita que un elemento clave no se incorpore al proceso de toma de decisiones.

1.6. Simulación con eventos discretos

A pesar de que los ensayos clínicos son generalmente la base de la evaluación económica, es muchas veces inviable su continuidad en el tiempo por lo que únicamente se dispone de resultados a corto plazo cuando se aplica esta metodología. Son aún más notorios los problemas que la aplicación de ensayos clínicos supone para la evaluación de programas de salud pública. En estos casos, los modelos matemáticos nos pueden ayudar a proyectar en el tiempo los resultados a corto plazo obtenidos mediante ensayos clínicos (Rodriguez-Barrios, 2008). Mediante esta metodología es posible expresar de forma matemática los elementos clínicos y epidemiológicos básicos de las enfermedades (Mar, 2010; Stahl, 2008). Los modelos de Markov son junto con los árboles de decisión los modelos más conocidos y utilizados en el ámbito sanitario para la evaluación económica de distintas intervenciones.

Los árboles de decisión son una de las primeras maneras formales de describir decisiones mediante una representación gráfica sencilla que no permiten retornar el flujo de los individuos (Stahl, 2008; Brennan, 2006). La representación gráfica consta de ramas que representan los eventos de una decisión y de nudos que representan los puntos de corte en los que se determinan los próximos eventos o bien se finaliza la secuencia de posibles eventos. Los árboles de decisión pueden recoger eventos recurrentes a lo largo del tiempo, pero tras pocas repeticiones el árbol puede tener demasiados nudos como para ser aceptable (Sonnenberg, 1993). A cada rama del árbol se le asigna una probabilidad la proporción de individuos en el modelo que optarán por esa rama (Fineberg, 1980). Cuanto más sencillo sea el árbol los datos requeridos serán más agregados lo cual facilita considerablemente la construcción de este tipo de modelos pero

aumentan a su vez las suposiciones sobre el comportamiento del sistema objeto de estudio (Davies, 2003). Los resultados se calculan mediante la valoración esperada obtenida mediante la suma de multiplicar la probabilidad por el número de entidades asignado de forma retroactiva (Stahl, 2008; Brennan, 2006).

Los modelos de Markov, en cambio, no tienen restricciones a la hora de representar gráficamente eventos recurrentes. Consideran el tiempo de forma discreta, dividido en ciclos, asumiendo que cada paciente podrá estar en un único estado a lo largo de cada ciclo. Las transiciones entre estados permiten que el individuo pueda volver al estado previo u pasar a uno nuevo al final de cada ciclo. La duración ideal del ciclo debería ser el intervalo de tiempo más corto clínicamente significativo (Mar, 2010). Lo habitual es que se utilicen ciclos anuales, pero según las características de la enfermedad representada también se pueden usar ciclos trimestrales o mensuales. En cada ciclo el paciente lleva a cabo una transición de un estado a otro en función de las probabilidades especificadas para el estado y el ciclo en que se encuentra. Estas probabilidades se denominan probabilidades de transición. Existen además estados absorbentes, como la muerte, de los que no hay transición posible, todos los individuos en este estado seguirán en el mismo estado con total probabilidad (Stahl, 2008). El modelo funciona hasta que todos los individuos están en el estado absorbente o hasta que se alcanza el horizonte temporal del estudio.

La principal limitación de los Modelos de Markov es la que se conoce como la asunción Markoviana o falta de memoria. Este concepto hace referencia a que las probabilidades de transición dependen únicamente del último estado en el que haya estado el paciente y no de su evolución. En la vida real los riesgos cambian con el tiempo, por ejemplo, las tasas de incidencia de cáncer mama aumentan con la edad. Los modelos resuelven este punto definiendo las probabilidades mediante fórmulas que incorporan la tasa en función de la edad, y modificando esta de forma automática en cada ciclo. Estas fórmulas permiten solventar el problema de forma puntual pero supone pensar en un modelo conceptual mucho más complejo que el necesario. Estos aspectos pueden ser fundamentales a la hora de aplicar este tipo de modelos para la representación de enfermedades complejas, como pueden ser las enfermedades crónicas, dado que al

intentar evitar estas limitaciones el modelo podría resultar en un sistema complejo y difícil de analizar (Soto-Alvarez, 2009).

Frente a las limitaciones que presentan los modelos previamente mencionados, los modelos de simulación con eventos discretos se están convirtiendo en la herramienta elegida dada su flexibilidad a la hora de representar sistemas complejos. La simulación es el conjunto de métodos y técnicas matemáticas para la imitación y reproducción de sistemas reales (Law, 2000). En la simulación con eventos discretos concretamente el tiempo en el sistema es continuo y sólo se ralentiza en los instantes en los que ocurre alguno de los eventos determinados para una mayor eficacia en el cálculo. Se podría decir que históricamente se distinguen dos formas de abordar el manejo del tiempo en los modelos: (1) avanzar hasta el siguiente evento o (2) avanzar un tiempo fijo cada vez. De hecho, el tipo (1) puede considerarse una generalización del tipo (2) por lo que cuando hablamos de modelos de simulación generalmente nos referimos a los modelos de simulación con eventos discretos (Law, 2000). No sólo en lo relativo al tiempo, que es tratada de forma continua, los modelos de simulación permiten una definir los eventos de interés o las probabilidades de ocurrencia en función de los atributos asignados de forma independiente a los individuos creados artificialmente. Este aspecto es esencial para la representación de poblaciones dinámicas y heterogéneas fundamental en la evaluación de políticas públicas.

En relación a los sistemas de salud, los modelos de simulación han sido utilizados generalmente para el análisis de sistemas de colas en un conjunto de servicios. El desarrollo de un modelo de simulación que reproduce su funcionamiento permite estimar los resultados que se hubieran obtenido en diferentes escenarios, así como conocer los errores que ser han producido durante el proceso (Azcarate, 2006). Sin embargo, su utilización no es habitual en la evaluación de intervenciones complejas en los que la historia natural de la enfermedad es parte fundamental del sistema a analizar. De hecho, actualmente la evaluación del programa de detección precoz de cáncer de mama se realiza en base a los indicadores de calidad establecidos en las guías europeas correspondientes (Perry, 2006).

En el desarrollo de un modelo de simulación con eventos discretos se distinguen tres pasos:

- Modelo conceptual: Se trata de la profundización en el conocimiento del tema de estudio. Es necesario sintetizar el sistema real y concretar los elementos fundamentales que se deben introducir en el análisis. Cada uno de los componentes (eventos, población, factores de riesgo) debe ser definido y se debe conocer a su vez la relación que existe entre los diferentes componentes así como el efecto de la intervención para cada caso. En este primer paso se decide cual será la población objetivo del estudio y el nivel de detalle que se incluirá en la evaluación. Es además el momento de concretar los resultados del modelo que permitirán responder a los objetivos del estudio.
- Estimación de los parámetros: Con el objetivo de poder trasladar el modelo conceptual a un modelo de computacional es necesario dar valores a los parámetros requeridos. Lo ideal en todo caso sería disponer de una base de datos de pacientes a nivel individual del cual poder estimar directamente el valor de cada uno de los parámetros necesarios. En la mayoría de los casos no se dispone de una única fuente de datos sino que es necesario combinar datos provenientes de bases de datos administrativas, resultados de ensayos clínicos o estudios previamente publicados.
- Desarrollo computacional: Una vez definidos el modelo conceptual y los parámetros necesarios para su implementación el último paso es la programación del modelo. Existen varios software comerciales diferentes destinados a la programación de modelos de simulación con eventos discretos como pueden ser Arena Rockwell Software o Simul8, pero también es posible utilizar lenguajes de programación generales para este propósito.

1.7. Evaluación del cribado mediante modelos matemáticos

En la evaluación de los programas poblacionales de detección precoz del cáncer más concretamente son tres las herramientas fundamentales para su evaluación: los registros de cáncer poblacionales, los ensayos clínicos y los modelos matemáticos.

Los registros de cáncer poblacional disponen de datos históricos que permiten analizar las tendencias de la incidencia, mortalidad y supervivencia del cáncer lo que permite ver el efecto del cribado en términos epidemiológicos. Sin embargo, el cambio en las tendencias no siempre es el efecto producido por el cribado dado que en ocasiones se pueden observar cambios de tendencia significativos también en países en los que el cribado poblacional no está implementado.

Como ocurre en todas las intervenciones clínicas, los ensayos clínicos aleatorizados son el diseño idóneo para la estimación no-sesgada del efecto de nuevas técnicas de cribado en la reducción de la mortalidad. No obstante, tal y como se ha mencionado previamente, existen diversas dificultades (éticas, económicas) en el desarrollo de este tipo de ensayos clínicos y su prolongación en el tiempo no es siempre suficiente para medir el efecto en la mortalidad.

Aun cuando se ha llevado a cabo algún ensayo clínico y los efectos del programa de cribado son conocidos, los modelos matemáticos basados en dichos resultados son necesarios para la estimación del efecto del programa en otros escenarios en cuanto a participación, incidencia, costes o características del test de cribado utilizado. De la misma forma, estos modelos aplicados para diferentes estrategias de cribado (edad de inicio y fin, intervalo entre invitaciones) ayudan a seleccionar la estrategia óptima a que maximice la función de beneficio a nivel poblacional (Habbema, Clin Lab Med 1982).

En Estados Unidos el Instituto Nacional del Cáncer (National Cancer Institute, NCI) financia y apoya un grupo de investigadores dedicados a la aplicación de la modelización estadística/matemática para el análisis del impacto de distintas intervenciones preventivas, como el cribado poblacional u otros tratamientos, en la incidencia y mortalidad del cáncer. Los modelos creados permiten proyectar tendencias futuras y por tanto ayudan en la determinación de las estrategias óptimas de control del cáncer. La red CISNET, creada en el año 2000, está constituido por seis grupos los cuales se dedican a cánceres con diferentes localizaciones: mama, cérvix, próstata, colon y recto, pulmón y esófago.

Uno de los principales objetivos del CISNET ha sido la creación un marco general de análisis que permitiera la comparación de los resultados de distintos modelos construidos

a nivel poblacional que a su vez ayudara en la toma de decisiones relacionadas con las política publicas. El marco establecido se centra en un modelo general del cáncer que después podrá ser modificado teniendo en cuanta una amplia gama de intervenciones. Los datos utilizados para la construcción del modelo podrán ser datos relativos a la demografía de la población como el género, la edad, la composición racial o la distribución de otros factores de riesgo en la población por una parte y por otra se incluirán los determinantes asociados a la intervención a estudio como nuevas modalidades de cribado o la difusión de tratamientos novedosos. Además, se tienen también acordados una serie de resultados tanto en términos de salud como de costes para todos los modelos construidos con el objetivo de que sean finalmente comparables.

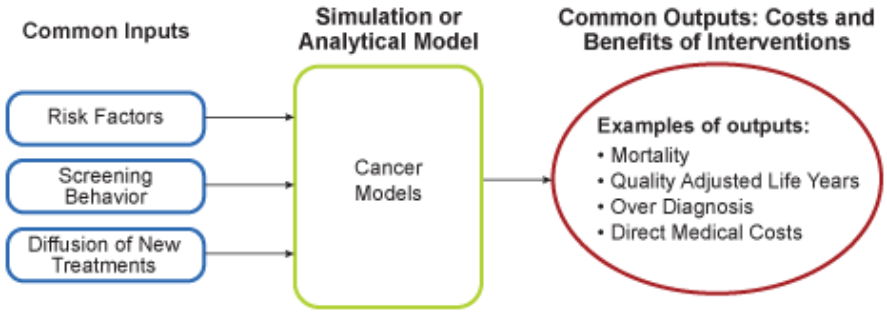


Figura 8: Marco general de los modelos del CISNET.

La mayoría de los modelos creados en el grupo CISNET son modelos de simulación, estos modelos permiten crear un número grande de personas artificiales y realizar un seguimiento de por vida que en un estudio observacional y experimental sería inviable tanto por el coste como por el tiempo necesario. La historia natural de un individuo incluye eventos como el nacimiento, la acumulación de los factores asociados al cáncer, la edad de inicio del desarrollo del cáncer en estado preclínico, la edad de progresión e inicio de los síntomas y la propagación de la metástasis, el diagnóstico del cáncer mediante cribado o de forma sintomática, el tratamiento y la muerte a causa del cáncer o por otras causas. Los modelos de simulación permiten además tener resultados a nivel

individual y por tanto asociar los factores de riesgo a los resultados en diferentes escenarios.

Puesto que la historia natural del cáncer en estado preclínico no es observable y no es éticamente correcto dejar sin tratar los cánceres detectados, la única forma de estimar la duración del estado preclínico es de forma indirecta. Cada modelo tiene sus propias asunciones en el desarrollo de la enfermedad en base a los datos de estudios de autopsias; ensayos de prevención, cribado o tratamientos; estudios de cohortes; registros de cánceres; u otros tipos de estudios. Cada estado del proceso de desarrollo se define por una serie de parámetros estimados a partir de datos empíricos o algoritmos estadísticos.

Este tipo de modelos incluye un amplio rango de cohortes de nacimiento y debe reproducir el cambio de los factores de riesgo, participación en el cribado o los tratamientos utilizados en cada momento y para cada edad a lo largo del tiempo de evaluación. En la evaluación de políticas de salud pública es necesario que la evaluación incluya las características de la población a nivel local. La distribución por sexo de la población o el proceso de envejecimiento de la población en los países desarrollados tienen un efecto que necesariamente debe tenerse en cuenta en el análisis. Los modelos poblacionales son aquellos sistemas dinámicos que combinan los efectos demográficos, epidemiológicos y clínicos a la hora de analizar el beneficio atribuible a una intervención. En consecuencia, los modelos poblacionales o multi-cohorte permiten tener en cuenta la heterogeneidad de la población (Ethgen and Standaert, 2012) y calculan el coste y el beneficio supuesto por el tratamiento a estudio en cada uno de los individuos que constituyen la población, aquellos que actualmente forman parte de la población diana y también aquellos que serán incluidos en dicha población en años futuros (Hoyle and Anderson, 2010). De esta forma los modelos que reproduzcan la dinámica poblacional permiten ver los resultados de distintas intervenciones en un mismo individuo y al mismo tiempo analizar su efecto en individuos con distintos factores de riesgo. De hecho son una gran ayuda para la evaluación de intervenciones poblacionales en un escenario en el que la interacción de las dinámicas poblacionales y su heterogeneidad, especialmente relacionada con el envejecimiento pueden determinar el resultado final de la evaluación (Ethgen and Standaert, 2012; Gold et al., 1996).

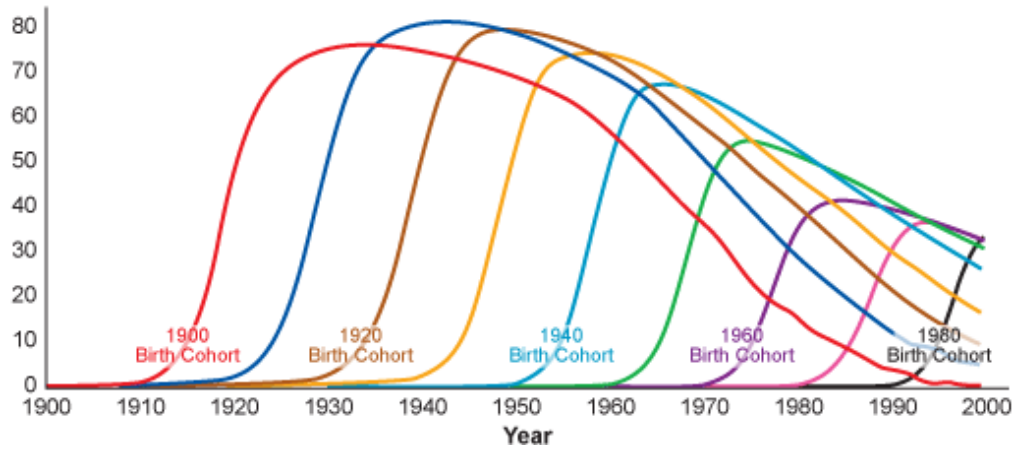


Figura 9: Diagrama de los modelo multi-cohorte o poblacionales

El cribado de cáncer de mama es una de sus principales líneas de investigación en la que trabajan principalmente en la medición del impacto del cribado poblacional tanto en la incidencia de cáncer de mama como en la mortalidad causada por la enfermedad. En el grupo que trabaja en los modelos de cáncer de mama está compuesto por siete grupos de investigación que trabajan de forma independiente:

1. Anderson Cancer Center, University of Texas, Houston, Estados Unidos.
2. University of Winsconsin, Madison, Estados Unidos.
3. Geogetown University Medical Center, Washington, Estados Unidos.
4. Erasmus University Medical Center, Rotterdam, Paises Bajos.
5. University of Rochester, Rochester, Estados Unidos.
6. Dana-Farber Cancer Institute, Boston, Estados Unidos.
7. Stanford University, Stanford, Estados Unidos.

Cada uno de los grupos puede escoger la metodología de modelización diferente pero el marco definido para la evaluación permite la comparación de los resultados obtenidos por todos ellos. De hecho, la selección del tipo de modelización es una de las asunciones en la construcción del modelo que pueden hacer variar los resultados finales. En el artículo de Boer et al. (Boer, 2004) se analizó el efecto de la metodología seleccionada en

la variabilidad de los resultados concluyendo que realmente la variabilidad viene dada por el conjunto de datos utilizado para la estimación de los parámetros necesarios en cada caso más que por la metodología de modelización seleccionada. Esto es, la variabilidad en los resultados sería mucho menor en caso de que todos los modelos utilizaran todos los datos disponibles para la estimación de los parámetros.

A nivel europea el grupo de la Erasmus University Medical Center de Rotterdam (Holanda) es sin duda el referente puesto que es el único grupo europeo incluido en el CISNET (Tan, van Oortmarssen et al. 2006). Este grupo aplica los modelos construidos dentro del grupo CISNET también a la evaluación del programa de cribado de cáncer de mama en los Paises Bajos. En la literatura también se pueden encontrar referencias a modelos basados en datos europeos (Arveux, Wait et al. 2003; Rojnik, Naversnik et al. 2008; Forastero, Zamora et al. 2010; Carles, Vilaprinyo et al. 2011; Brailsford, Harper et al. 2012) pero ninguno es un modelo que evalúe el efecto del cribado de cáncer de mama en una población real dinámica que incluya en cada momento las cohortes correspondientes.

1.8. Predicción del riesgo individual de cáncer de mama.

Actualmente, la edad y el sexo son los únicos criterios de definición de la población diana objeto de cribado. Sin embargo, se ha descrito que la edad al nacimiento del primer hijo, los antecedentes familiares, la densidad mamográfica o los factores genéticos, también se asocian a un mayor riesgo (El-Bastawissi, 2000; Titus-Ernstoff, 2006). Disponer de una estimación fiable del riesgo individual de cáncer de mama a partir de factores conocidos podría ayudar a personalizar del cribado del cáncer y una optimizar el uso de recursos a nivel poblacional.

La consideración del riesgo individual de cáncer de mama en la definición de nuevas estrategias de cribado es novedosa. Los modelos matemáticos desarrollados por el CISNET han sido utilizados también para la investigación de estrategias de cribado

óptimas, considerando por ejemplo distintas edades para el comienzo y el final del cribado (Feuer, 2006; Fryback, 2006; Lee, 2006). Recientemente, mediante un estudio de coste-efectividad, Schousboe et al. (Schousboe, 2011) han propuesto diferentes periodicidades entre las mamografías de cribado, según el riesgo de cáncer de mama medido en función de la densidad mamaria, la historia familiar y las biopsias previas.

La estimación del riesgo de cáncer de mama ha sido motivo de publicación de diversos trabajos en las últimas décadas. El modelo creado por Gail et al. en el año 1989 es, sin duda, el más conocido y utilizado hasta el momento en la predicción del riesgo de cáncer de mama (Gail, 1989). Este modelo ha sido además el referente para nuevos modelos que se han desarrollado posteriormente. La modificación del modelo de Gail, mediante la inclusión de la densidad mamaria y el peso de la mujer como factores de riesgo, dio lugar al modelo desarrollado por Chen et al. (Chen, 2006). El modelo descrito en el año 2006 por Barlow et al. (Barlow, 2006) consideró además de la densidad mamaria, factores como la terapia hormonal sustitutiva, el índice de masa corporal, la raza o la etnia, que hasta aquel momento no habían sido incorporados como variables predictivas. Existen a su vez otros modelos que también tienen en cuenta el riesgo de ser portador de los genes BRCA1 o BRCA2. Modelos como el BRCAPRO (Parmigiani, 1998) o el Tyrer-Cuzik (Tyrer, 2004) se centran básicamente en los antecedentes familiares de cáncer de mama y ovario.

La generalización del uso de estos modelos requiere que sean validados previamente en poblaciones distintas a las que fueron desarrollados, dadas las posibles diferencias en la distribución de los factores de riesgo y en la epidemiología del cáncer de mama. Una vez considerada su validez externa, se puede plantear su utilización en la estimación del riesgo individual y la personalización de estrategias de cribado. La eficiencia de los programas de cribado mejoraría si se dispusiera de estimaciones fiables del riesgo individual de cáncer de mama ya que permitiría personalizar el protocolo del cribado. El objetivo principal de este estudio fue evaluar la calibración y el poder de discriminación de modelos de predicción del riesgo individual de cáncer de mama, sin información genética, en una cohorte de mujeres en un programa de detección precoz en España.

2. Objectives

MAIN OBJECTIVE:

To assess the epidemiological and economic impact of the breast cancer early detection programme in the Basque Country.

SECONDARY OBJECTIVES:

- To meassure the effect of the screening programme in breast cancer incidence and mortality.
- To carry out cost-effectiveness analysis of a population level screening programme versus an unscreened population.
- To performe budget impact analysis of the screening programme.
- To calculate the discriminatory power of published individual risk prediction models.

3. Material and methods

3.1. Screening evaluation through a simulation model.

A discrete event simulation model (Karnon, 2012; Stahl, 2008) was built to reproduce the natural history of breast cancer (BC) for women invited to participate in the programme and the characteristics of the women invited to the breast cancer screening programme in the Basque Country (BCSPBC) since its beginning in 1996 through 2011.

The screening programme comprised multiple cohorts of women; a cohort was defined as a group of women invited to participate for the first time in the screening programme in a calendar year. Following the terminology used by Hoyle and Anderson (Figure 10) the cohort starting screening in the current year is defined as the incident cohort, and those already undergoing screening from previous years are known as the prevalent cohorts (Hoyle, 2010). These terms do not correspond to any disease state in this context but to screening eligibility.

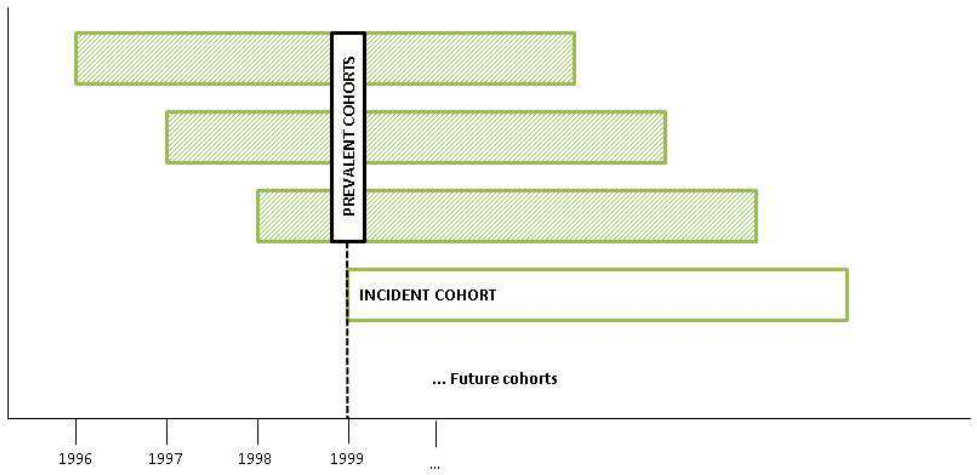


Figure 10: Incident and prevalent cohorts in multi-cohort model.

Our model reproduced the entire female population invited into the programme during the period 1996–2011. In 1996, during the first year of implementation, all the women invited belong to a unique incident cohort of women aged 50 to 64 years. In subsequent

years, instead, the incident cohort included those aged 50 to 51 years. Actually, the target population also included several prevalent cohorts, apart from the incident cohort.

The evaluation period was defined as 1996 through December 31, 2011, as the target population of the programme was changed during 2012 and extended to women in their 40's with a first-degree family history of BC. It is considered that women with first-degree family history of BC have higher lifetime risk to develop BC, therefore this new extension of the programme was not included in the evaluation of the BCSPBC this time.

Exactly 414,041 life histories were created, one for every woman invited at least once into the BCSPBC from 1996 through 2011. A simplified diagram of the model is shown in Figure 11 developed using Arena Simulation Software (Versión 14.0, Rockwell software, Milwaukee, WI).

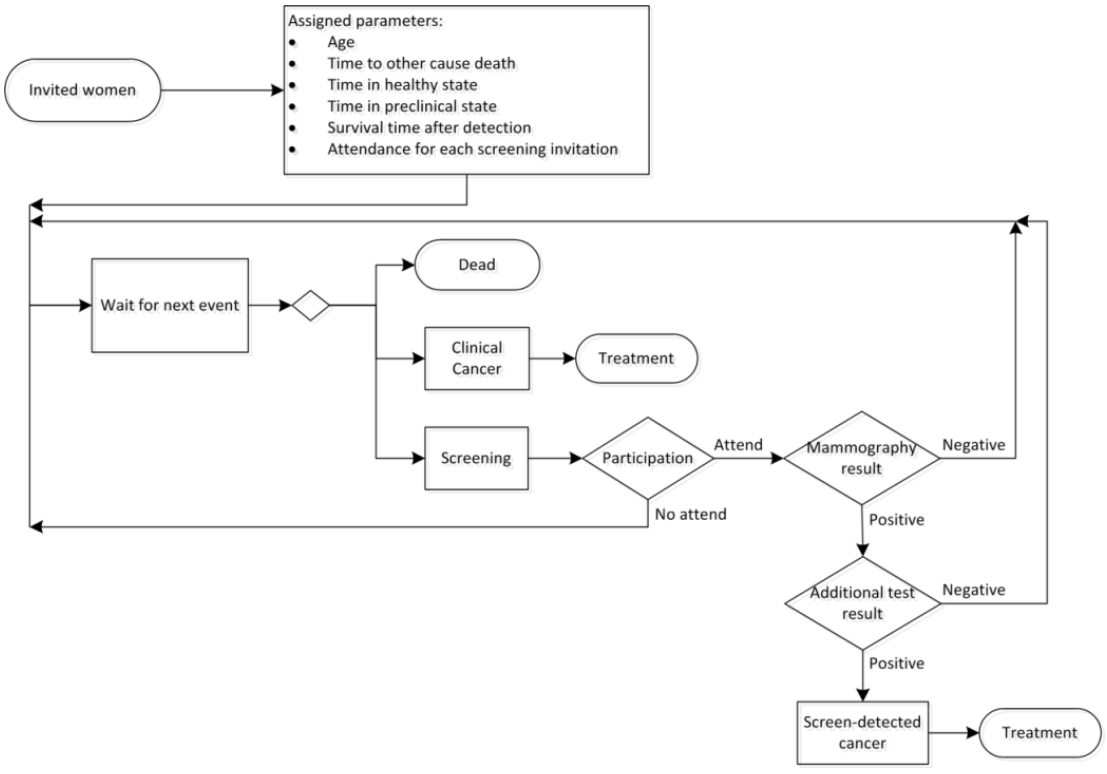


Figure 11: Conceptual model

Table 7 shows data sources used to obtain the parameters necessary for the implementation of the model.

Table 7: Model input and validation parameters.

Input data	Source
Invited population	
Number of women invited for the first time	Screening programme data
Age distribution	Screening programme data
Participation rate	Screening programme data
Time until event	
Other cause mortality	Basque mortality registry
Breast cancer mortality	Basque mortality registry
Time till pre-clinical state	Rue et al, 2009
Pre-clinical state duration	Lee and Zelen, 2006
Age- and stage-specific breast cancer survival	Vilaprinyo et al.
Detection data	
Clinically detected cancer stage distribution	Basque cancer registry
Programme sensitivity and specificity	Screening programme data
Screen detected cancer stage distribution	Screening programme data

Validation data	Source
Invited population	
Total number of invited women	Screening programme data
Total number of mammograms	Screening programme data
Recall rate	Screening programme data
Remitted for additional test	Screening programme data
Detection data	
Age- and year-specific breast cancer incidence	Basque cancer registry
Screening-detected cancers	Screening programme data

The exact number of women invited to participate in the programme for the first time and their ages at that time were available from the programme database (Table 8). Precisely, 414,041 women were invited to the programme and 320,366 attended, 78% of the invited population. The age distribution of the invited women changed during the study period. From 1996 to 1998 during the programme implementation, the population consisted only of the incident cohorts aged 50 through 64 years. In subsequent years, instead, the incident cohort included those aged 50 to 51 years. Actually, the target population also included several prevalent cohorts, apart from the incident cohort. The extension of the target population from 50 to 64 years and then 50 to 69 years began in 2006, with women aged 65 years continuing in the programme until age 69. The screening test for BCSPBC consisted of mammography with double projection (cranio-caudal and oblique lateral view) carried out biennially (Sarriugarte et al, 2011).

Table 8: Number of women invited into the BCSPBC and participation rates (%)..

	First invitations		Successive invitation	
Year	Number of women	Participation	Number of women	Participation
1996	7,835	79.71	0	-
1997	67,719	72.94	0	-
1998	87,967	78.26	16,702	71.49
1999	41,841	84.60	51,037	64.57
2000	17,426	96.27	80,399	74.77
2001	18,902	90.45	86,792	70.82
2002	16,401	90.04	86,110	74.54
2003	21,109	84.38	87,877	74.59
2004	16,363	87.26	86,327	75.08
2005	14,043	89.49	91,996	75.35
2006	16,804	86.39	114,691	73.97
2007	17,018	87.92	105,850	75.18
2008	17,847	83.85	110,542	75.15
2009	18,510	85.68	116,330	75.51
2010	17,711	88.48	120,481	79.45
2011	16,545	91.21	128,836	79.49

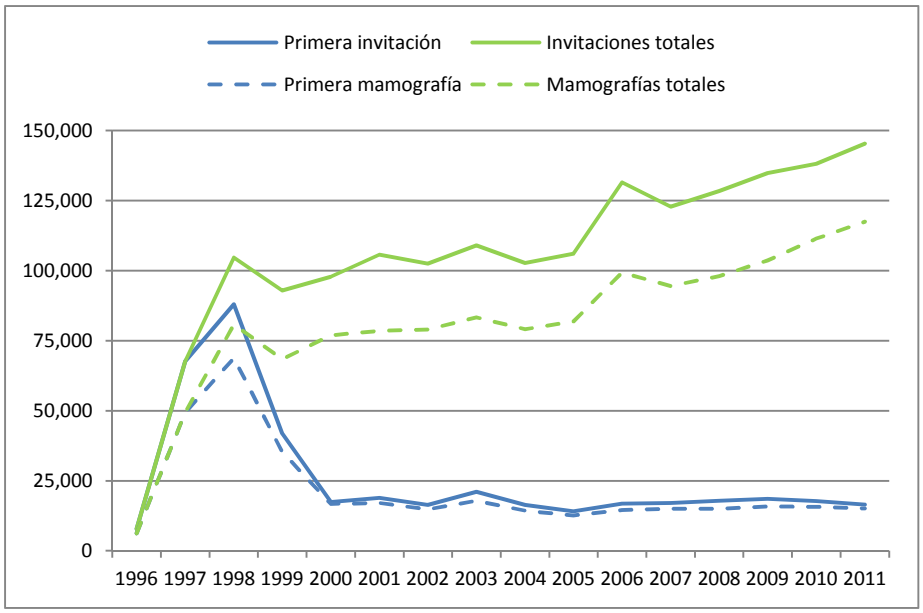


Figure 12: Invited and screened women for each calendar year.

Mortality from causes other than BC was randomly assigned, depending on the woman's birth cohort, based on an empirical function. All-cause and BC-caused mortality data were obtained from the Basque mortality registry for the period 1986–2010. The high quality of the Basque cancer registry data has been demonstrated by Izarzugaza et al. (Izarzugaza, 2010) The Basque Statistics Institute (EUSTAT) provided the population of Basque women by age and birth cohort. We applied an actuarial method that removes breast cancer as a cause of death to estimate the age at death from causes other than BC, by birth cohort (Table 9) (Vilaprinyo et al, 2008).

Table 9: Proportion of survivors, lτ, and proportion of survivors after removing breast cancer as a cause of death, l-bc, by age and cohort of birth.

	Birth cohort								
	1930-1934			1940-1944			1950		
	l^{τ}	l^{-bc}	Dif	l^{τ}	l^{-bc}	Dif	l^{τ}	l^{-bc}	Dif
20-24	0.9981	0.9989	7.5	0.9984	0.9990	6.6	0.9986	0.9992	5.43
25-29	0.9972	0.9981	9.4	0.9975	0.9984	8.2	0.9980	0.9986	6.76
30-34	0.9961	0.9972	11.2	0.9966	0.9976	9.8	0.9972	0.9980	8.08
35-39	0.9948	0.9962	13.5	0.9955	0.9967	11.8	0.9963	0.9972	9.77
40-44	0.9934	0.9951	17.1	0.9942	0.9957	15.0	0.9952	0.9965	12.41
45-49	0.9916	0.9939	22.8	0.9926	0.9946	20.1	0.9939	0.9956	16.58
50-54	0.9893	0.9924	31.5	0.9906	0.9934	27.7	0.9923	0.9945	22.94
55-59	0.9861	0.9906	44.6	0.9878	0.9918	39.3	0.9900	0.9932	32.48
60-64	0.9814	0.9879	64.8	0.9837	0.9894	57.1	-	-	-
65-69	0.9739	0.9837	97.9	0.9771	0.9857	86.2	-	-	-
70-74	0.9612	0.9767	155.5	-	-	-	-	-	-
75-79	0.9387	0.9646	259.3	-	-	-	-	-	-
80-84	-	-	-	-	-	-	-	-	-
>85	-	-	-	-	-	-	-	-	-

When the model for the screened scenario was calibrated and validated the same model was run also for the unscreened scenario involving the same female population invited at least once into the BCSPBC during the study period. All the created entities were cloned to obtain two identical populations (screened and unscreened) in each run. No ethical

approval or consent was required as no experimental research on humans was involved in this study. However, the Ethics Committee for Clinical Research in Gipuzkoa Health Area evaluated and approved the study.

3.2. Natural history of breast cancer.

We modelled the natural history of BC using the approach adopted by Lee et al. (Lee, 2006). Four main states of health were distinguished: (1) disease-free or undetectable BC; (2) asymptomatic BC that can be diagnosed by screening or preclinical phase; (3) symptomatic BC diagnosed clinically; and (4) death from BC.

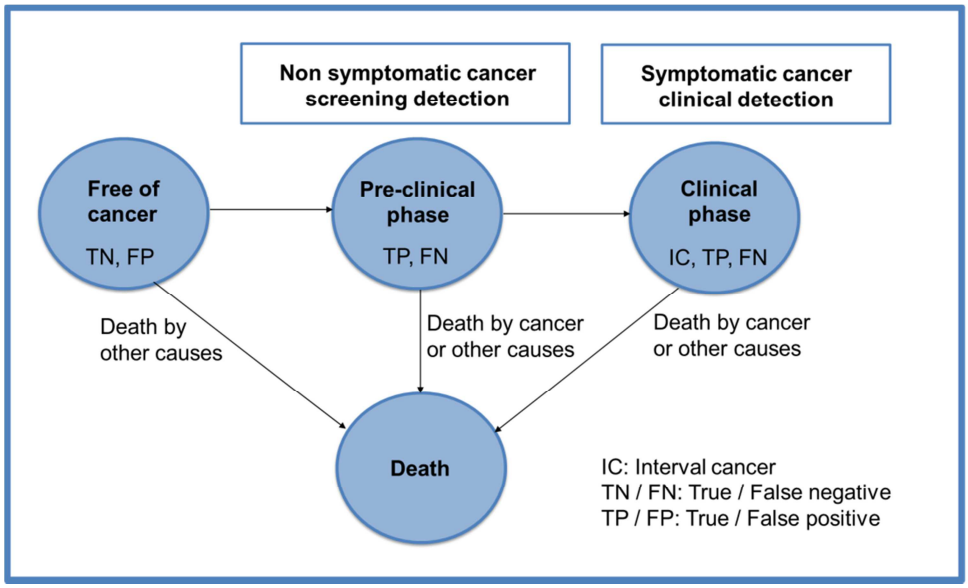


Figure 13: Natural history of breast cancer

The age distribution used to assign the onset of the preclinical phase was obtained from Rue et al. (Rue, 2009). On the basis of BC incidence from Catalan cancer registries and a distribution of the sojourn time or duration in the pre-clinical state, those authors used a

generalized linear model with a Poisson distribution and a polynomial parameterization for the variables of age and cohort for the estimation of BC incidence when no data was available (Rue, 2009). Cohort effects enabled including upward breast cancer incidence trends in our model.

$$t_{s \to p} \sim F(t) = 1 - S(t) = 1 - \exp\left(- \int_0^t tasa_{\ cohorte}\ (u)\, du \right)$$

Due to the reduced volume of the Basque population figures in the Basque cancer registry were not enough in order to reproduce the procedure applied in the study carried out in Catalonia. Therefore, we decided to compare BC incidence in the Basque Country and in Catalonia with the aim of calibrating the functions available (Figure 14) and obtain the best fitting possible for the observed incidence in the Basque Country during the evaluated period.

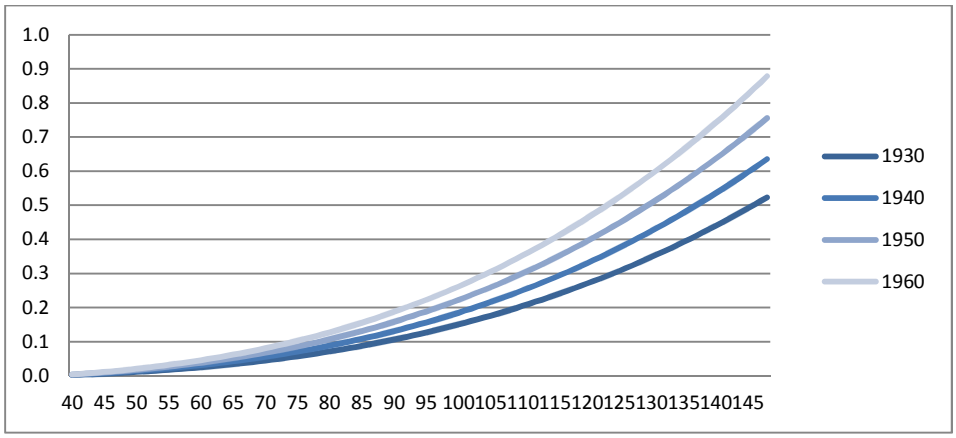


Figure 14: Cumulated probability of breast cancer onset by age and birth cohort.

We assumed, as did Lee et al., that the sojourn time of the pre-clinical phase follows an exponential distribution as based on results of clinical trials (Lee, 2006). The mean value

used by Lee et al. for women aged 50 years or more was 4 years. In the model built for the Basque programme evaluation this value was calibrated in order to reproduce the observed BC incidence adjusted by age.

One of the main assumptions in this model was that every woman who reached the clinical state would be diagnosed clinically at the beginning of this state. Thus, we used the age-specific distribution of BC detection stages observed in the cancer registries of the Basque Country in 1995, before the screening programme began for clinically detected BC (Table 10). In situ carcinomas are also included in the model as the lowest stage in which BC could be detected.

Table 10: Distribution of breast cancer detection stages.

Detection stage	In situ	Stage I	Stage IIa	Stage IIb	Stage III	Stage IV
Clinically detected cancer						
50-59	10.00	32.63	24.75	15.75	9.00	7.88
60-69	7.42	21.72	22.86	26.29	13.72	8.00
>69	4.35	12.11	27.85	12.11	24.22	19.37
Screen detected cancer	In situ	Stage I	Stage IIa	Stage IIb	Stage III	Stage IV
Period 1996-1999						
50-59	19.69	49.71	19.30	7.60	3.12	0.58
60-69	17.94	50.93	19.18	6.80	4.12	1.03
Period 2000-2005						
50-59	18.77	49.16	20.34	6.39	4.65	0.70
60-69	18.08	57.18	15.60	5.33	3.43	0.38
Period 2006-2008						
50-59	18.77	49.16	20.34	6.39	4.65	0.70
60-69	18.08	57.18	15.60	5.33	3.43	0.38
Period 2009-2011						
50-59	18.76	49.47	21.18	4.99	4.54	1.06
60-69	15.55	54.50	19.97	5.56	3.93	0.49

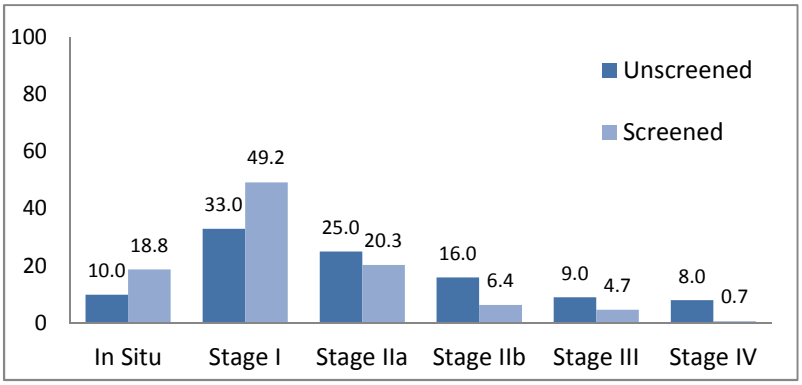


Figure 15: Distribution of the breast cancer detection stages.

On the basis of the work by Vilaprinyó et al. (Vilaprinyo, 2009), we applied distributions of age and stage-specific survival in women diagnosed either clinically or by screening. We assume that the main effect of the screening programme is based on the early diagnosis of the cancer. Thus, when screening is applied less cancers are detected in advanced stages, but once they are diagnosed cancer prognosis depend on the detection stage and no on the type of detection (by screening or simptomatic). In the simulation model each diagnosed woman was assigned two ages at death and the minimum of these two ages determined the cause and age of death for each woman.

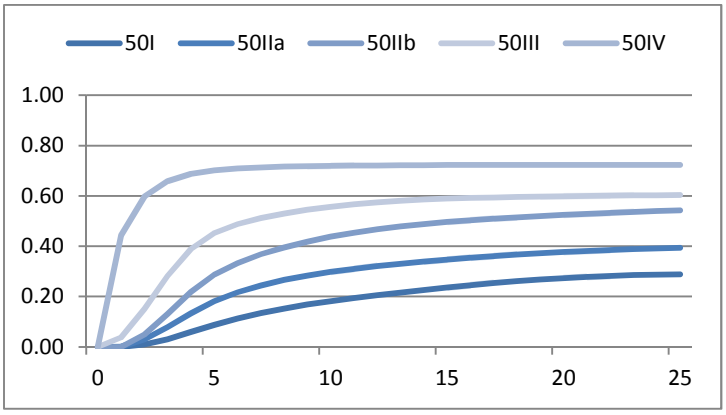


Figure 16: Survival functions by age and detection stage.

All these parameters which define the natural history of BC are determined at the beginning of the modelling, during the definition of the entities which represent each of the women on the target population. Afterwards, each entity of the simulated population was cloned to obtain two identical populations in order to compare the screened and unscreened scenarios.

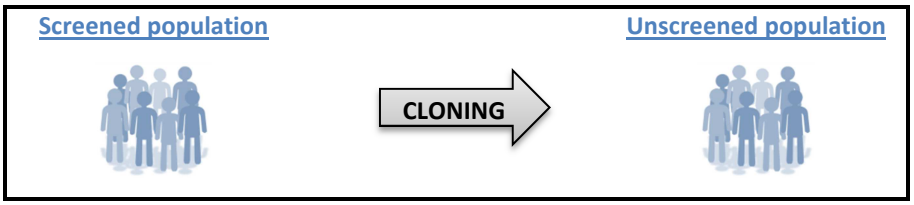


Figure 17: Graphical representation of the population cloning procedure.

3.3. Input data

The good quality of the programme data base allowed to calculate the exact number of women invited for the first time into the BCSPBC from 1996 through 2011, exactly 414,041 women. Their age distribution was also obtained from the programme data base. From 1996 to 1998 during the programme implementation, the population consisted only of women invited for the first time, that is, cohorts aged 50 through 64 years. In subsequent years, instead, only cohort aged 50 to 51 years were invited for the first time. Actually, the target population also included several cohorts that had previously been invited to participate in the programme, apart from those that received the invitation for the first time. The extension of the target population from 50 to 64 years and then 50 to 69 years began in 2006, with women aged 65 years continuing in the programme until age 69 (Sarriugarte, 2011).

Theoretically women were biennially invited to the screening programme, however, during the programme implementation, time interval between invitations was not exactly

two years. The simulation model should reproduce the total number of invitations annually carried out by the programme based on the observed number of first invitations delivered and indicated time interval between invitations. As using a constant interval of two years the simulated invitations did not fit well observed data, we assumed that the interval between invitations should be maintained constant and its value estimated using calibration in order to reproduce the observed number of annual invitations in the screening programme.

The total number of mammograms performed in the programme was determined by the number of invited women (including early recalls) and annual attendance rates, which were exactly known from the programme data base (Table S3). Annual attendance rates were considered independent as correlation of the participation in first and repeated screening rounds was not available.

Sensitivity and specificity of the entire early detection programme, as well as distribution of breast cancer stage for cancers detected by screening, varied during the study period. Four phases were distinguished during the studied period due to the variability of sensitivity and specificity values and screen-detected BC stage distribution: (1) from 1996 to 1999, the implementation phase, when most of the women invited to the programme received their first invitation; (2) from 2000 to 2005, the prevalence phase, when the percentage of women invited for the first time was much lower than the percentage of women invited for successive mammograms; (3) from 2006 to 2008, extension phase, when the programme was extended to women aged 65 to 69 years; (4) from 2009 to 2011, digital phase, when the switch to digital mammography occurred.

Observed screening mammography results were used together with the number of invited women and number of screening-detected breast cancers and observed interval cancers (Table 11) to calculate sensitivity and specificity for each of the defined phases.

$$Sensitivity = P(+|C) = \frac{Detected\ cancers}{Detected\ cancers + Interval\ cancers}$$

$$Specificity = P(-|\bar{C}) = \frac{Negative\ results - Interval\ cancers}{Negative\ results - Interval\ cancers + False\ positives}$$

Where C represent women in pre-clinical or clinical state and $\bar{C}$ those that are healthy.

Table 11: Sensitivity and specificity of the BCSPBC.

Year	1996-1999	2000-2005	2006-2008	2009-2011
Total screened women	207063	400539	480192	333877
Detected cancers	1012	1407	1870	1366
False positives	19687	37455	30246	19509
Interval cancers (IntBC)	51	280	369	225
True negatives (Negatives – IntBC)	186313	361397	447707	312777
Sensitivity	95.20	83.40	83.52	85.86
Specificity	90.44	90.61	93.67	94.13

In the model, a positive or negative screening result was assigned based on the presence or absence of BC in each woman and the corresponding sensitivity and specificity of the programme. That is, the probability for a negative test in healthy women would be determined by the specificity of the programme for that year, whereas in the case of women that had already developed BC, sensitivity would define the probability for a positive result. Based on these probabilities random numbers were used to stablish the final result for each woman.

Distributions of disease stages for screening-detected cases were also obtained for the different phases of BCSPBC (implementation, prevalence, extension and digital) with use of observed data from the screening programme (Table 10). These figures were in line with other stage distributions used in similar studies carried out in different countries (Lee, 2006; Vilaprinyo, 2014). In addition, as two identical populations were created for the comparison of the screening and no-screening scenarios, the same random numbers were used to simulate the stage distribution for the clinically and the screening-detected cancers in the same woman, in order to estimate the advance in detection stage due to screening (Figure 18).

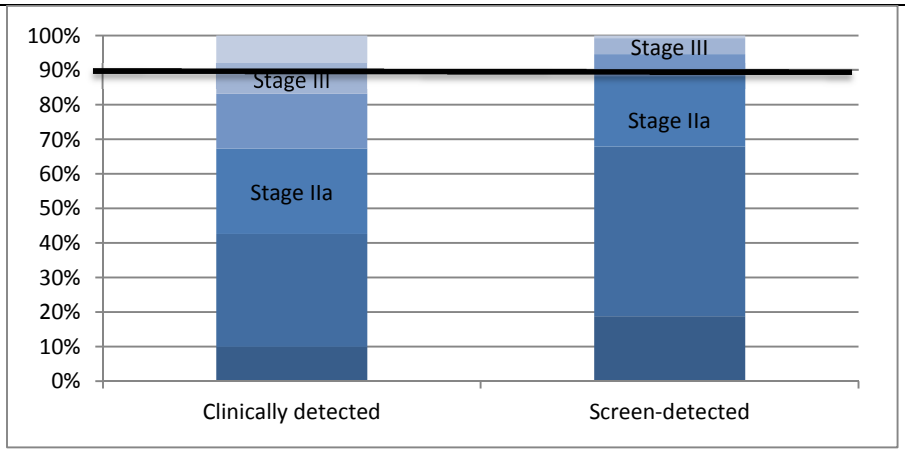


Figure 18: Methodology applied for cloning the population taking into account early detection.

3.4. Calibration of the non-observable parameters

The most common use of calibration is to estimate unobservable model parameters by only allowing these parameters to vary in the calibration process. Moreover, even when parameters have been observed directly, these parameters may have different levels of precision, leading some to advocate that all natural history and other relevant parameters in the model (unobservable and observable) should be allowed to vary in the calibration process. The comprehensive inclusion of parameters facilitates the representation of correlation between input parameters, and permits the testing and adjustment of the global consistency of the model. However, it does not exclude the need to investigate and to represent the correlation in the model.

Karnon y Vanni (Karnon, 2011; Vanni, 2011) described the methodology chosen for our simulation model calibration. They categorized the calibration process into the following seven steps:

1. Select the parameters that should be varied in the calibration process.
2. Calibration targets.
3. Mesure of goodness of fit.

4. Parameter search strategy.
5. Convergence criteria.
6. Calibration process stopping rute
7. Integrate in the model calibration results.

The model was run in the screened scenario for the whole female population invited at least once into the BCSPBC during the study period in order to reproduce the actual performance of the programme. Three main parameters were selected to vary in the calibration procedure:

- Time between consecutive invitations
- Age distribution of preclinical phase onset
- Mean duration of preclinical phase.

First, we calibrated the time between invitations considering that it was not influenced by other unobserved parameters. Afterwards, we calibrated jointly two factors. The first one was the relative risk (RR) for the incidence function and the second one the mean value for the preclinical state duration which prior estimate was 4.0.

The selection of the calibration targets is another important step in the calibration process. There are no exact criteria to choose the calibration targets that are necessary to the process. However, since there are regional differences in disease epidemiology and management pathways, local data should be preferentially used as targets. It is important that a model accurately represents the condition of the population for which the decision is being made and from which empirical data were obtained to use as targets. In the calibration process of the simulation model that should reproduce the implementation of the BCSPBC during the period 1996-2011, the following available data were selected as calibration targets:

- Total number of invitations by calendar year.
- Total number of breast cancers detected by calendar year.
- Total number of screen-detected breast cancers by calendar year.

In the next step goodness of fit measures are used to evaluate how close the model predictions are to target data. In the statistics literature, the most commonly used measures of goodness of fit are least squares, chi-squared (or wheighted least squares) and the likelihood (Cooper, 2007).

Finally, the goodness-of-fit measure applied in our model to assess the difference between observed and estimated outcomes (mean value of 25 simulations) was the chi-square statistic.

$$\chi^2 = \sum_a \left(\frac{y(a) - f(a|\theta)}{\sigma_a} \right)^2$$

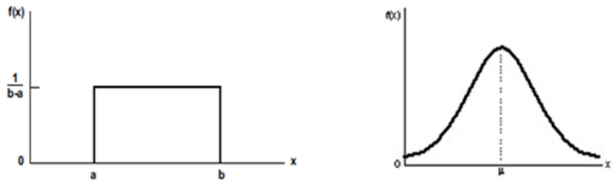
The overall chi-square statistic of each hypothesis was calculated as the sum of the chi-square statistics calculated for the analysed years. We assumed outcomes for each year to be independent and uncorrelated. Finally, we included in the model the parameter value for which the overall chi-square was the lowest (convergence criteria).

In the following step the term “parameter search strategy” refers to the method used to search for parameter values or sets of values that produce model outputs that match specified calibration targets most closely. Is therefore an optimization process of finding the conditions that minimize the difference between the estimated and observed outcomes. Unfortunately, there is no perfect optimization algorithm and so it is necessary to consider the most appropriate methods and even try more than one method, or combiantions of methods, in order to achieve the optimum values for the calibrated parameters. Consequently a combination of random search and grid search algorithms was selected as search strategy. Each of this search strategies have pros and cons that were profitable in this case.

In a random search method, distributions are assigned to each parameter in the model and multiple sets of parameter values are sampled using a random number generator.

Each set is then used in the model and theGOFis calculated. The set (or sets) that results in the optimum GOF result(s) is selected according to the convergence criteria.

The main advantages of the random search strategy are that it is intuitive and relatively easy to programme. The main disadvantage is that random searching is not efficient in covering the entire parameter space. With a random search strategy, increasing numbers of searches improves the chance that the global extremum has been identified, but we cannot be certain that the extremumidentified is global and not local.



$UNIFORM(a, b)$ $NORMAL(\mu, \sigma)$

Figure 19: Probability density functions.

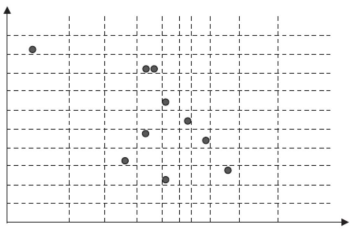


Figure 20: Sample elements selected using random seach algorithm.

In a grid search method, the parameter search takes place across the different possible combinations of parameter values. Conceptually, if just two parameters (X1 and X2) were varied in the model, the space could be represented in two dimensions. By considering this two dimensional space as in Figure 21, it is simple to understand how the grid parameter searchmethod works. For example, if the lower and upper bounds of the two variables are li and ui (i = 1, 2), for simplicity we could divide the ranges into two equal parts, with three considered values per parameter (vi = 3). This method involves setting

up a suitable grid in the parameter space, evaluating the GOF estimate at all the grid points (nine points in the example), and finding the grid point that best minimizes the GOF.With each additional parameter, the number of dimensions required to represent the space also increases accordingly and, in most practical problems, the grid search method requires prohibitively large numbers of model evaluations.

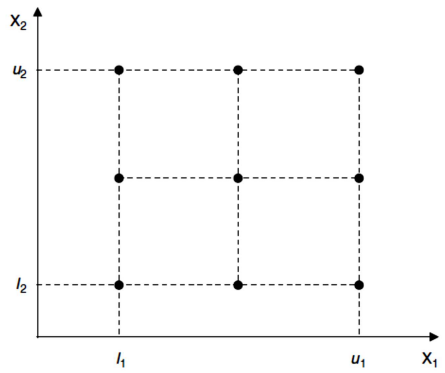


Figure 21: Sample elements selected using grid search algorithm.

Convergence criteria, acceptance criteria and the acceptance threshold are terms that describe the process of defining acceptable sets of input parameter values. We will assume that the parameter value that best minimizes (or maximizes, depending on the GOF measure used) the GOF estimate, this is the acceptance criterion, will be the optimal parameter value to be used in the model.

First, we calibrated the time between intervals considering that it was not influenced by other unobserved parameters. At the beginning, we used a random search algorithm considering different values from a normal distribution centred in 2 years and standard deviation 0.5. Based on these results, we continued using a grid search algorithm, running 25 simulations for 10 different values between 2.11 and 2.20. The goodness-of-fit measure applied to assess the difference between observed and expected outcomes was the chi-square statistic. We included in the model the parameter value for which the

overall chi-square statistic was the minimum: 2.18 year between consecutive invitations (Figure 22).

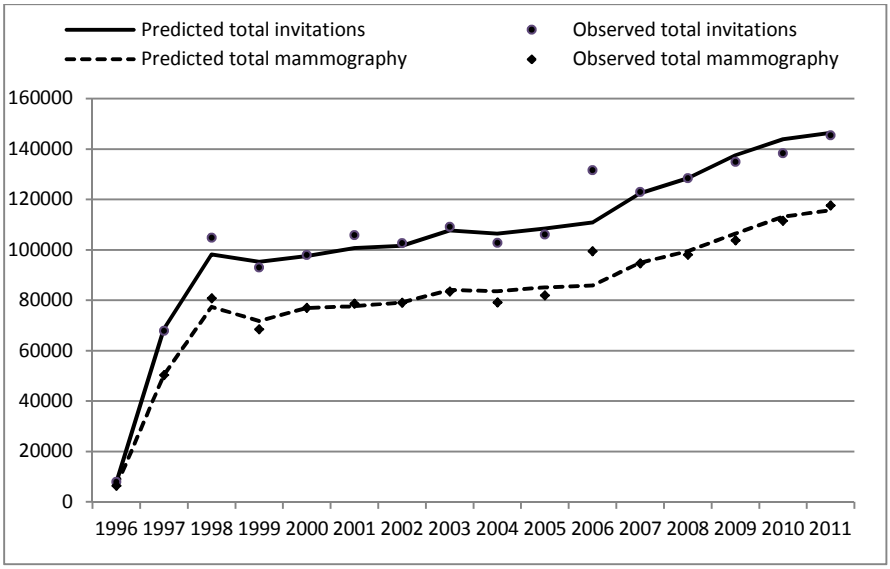


Figure 22: Total number of women invited to join the programme and the number of the mammograms carried out.

Afterwards, we calibrated jointly two factors. The first one will be the relative risk (RR) for the incidence function. The second multiplier will be used to calibrate the mean value for the preclinical state duration which prior estimate was 4.0. Thus we will calibrate the factor t to obtain a final mean preclinical state duration 4t. We considered as target outputs the number of screening-detected cancers from 1996-2011, together with total cancer detection rates by age group (50-54, 55-59, 60-64, 65-69) for the period 1999-2009. Random search algorithm was used also in this case considering Normal(1,0.25) distribution for both parameters for a first approximation and a grid search algorithm centred in $0.87 \leq RR \leq 0.90$ and $0.85 \leq t \leq 0.90$. The goodness-of-fit measure used in this case was also the chi-square statistic. The final relative risk used for BC incidence functions was 0.88, and the mean time in preclinical state 3.44 years (Figure 23).

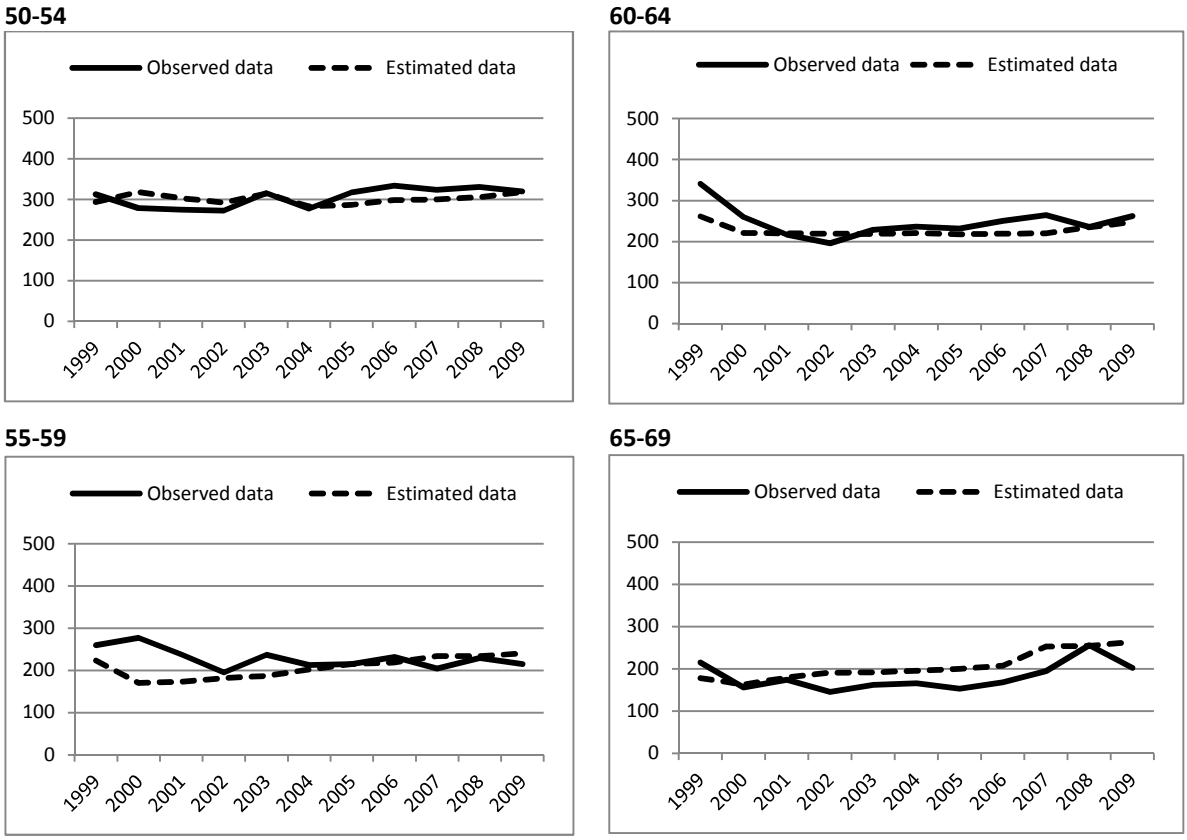


Figure 23: Breast cancer incidence calibration by age group and calendar year.

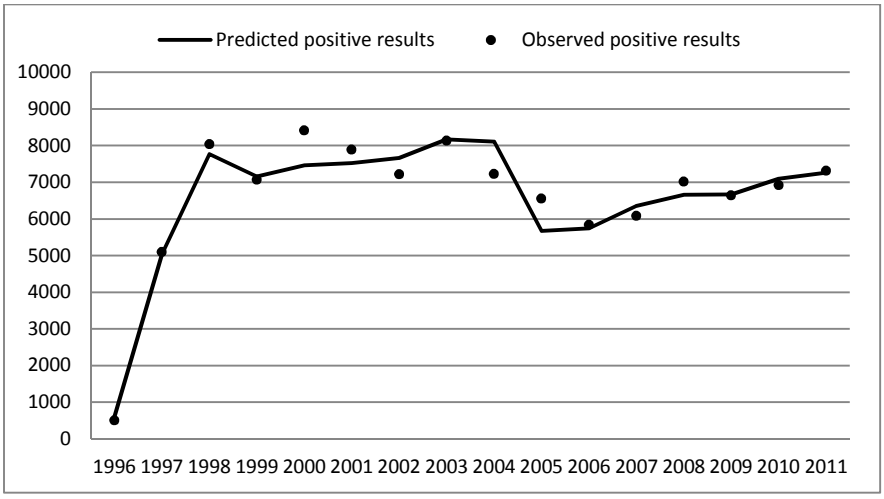


Figure 24: Total number of positive mammogram results in the breast cancer screening programme.

For model validation, we compared the estimated results for the screened population (multi-cohort model) with the observed indicators from BCSPBC and the Basque cancer registries such as number of invited women (Figure 22), number of mammograms carried out in the programme (Figure 22), age-specific breast cancer incidence (Figure 23) or the number of women with a positive mammography result (Figure 24). We also confirmed that life expectancy for women from the general population and women who died from BC was concordant with the observed data (Table 12).

Table 12: Validation of the mean life expectancy for women in the general population and median survival time corrected by lead time for women with death from BC.

	Theoretical	Estimated
General population	83.70	82.61
BC death survival*		
Stage I	9.03	6.34
Stage IIa	6.46	4.77
Stage IIb	5.14	4.19
Stage III	3.41	2.74
Stage IV	0.80	0.63

*Median BC survival times when no other cause deaths occur are shown as theoretical. Estimated median survival times for BC deaths are lower than theoretical as women with greater BC survival time die from other causes.

3.5. Outcomes assessment

When the model for the screened scenario was calibrated and validated the same model was run also for the unscreened scenario involving the same female population invited at least once into the BCSPBC during the study period. All the created entities were cloned to obtain two identical populations (screened and unscreened) in each run.

We first used a multi-cohort model that allowed the best approach to reproduce population dynamics in the Basque Country as well as the natural history of breast cancer. However, the assessment of the balance between benefits and harms from preventive programmes typically requires a long follow-up thus the key for its

interpretation is to achieve the steady state that is defined as the time when each recently observed behaviour of the system will remain constant in the future (Asmussen, 2007). As the first 15 years of the BCSPBC evaluated in this study were not enough to achieve this state, we cannot ensure that the differences between the two scenarios estimated in 2011 will remain in the future.

To further understand the effects of a programme that requires long-term follow-up, we reran the model using a single cohort of 50,000 women aged 50 years who were invited into the programme for the first time in 1996, assuming 100 % participation and life-time follow-up (Ormiston-Smith, 2013). The inputs used to extrapolate the results from 2011 on were based on the same parameters as those that were used for 2011.

In order to include variability of population characteristics in the model, the multi-cohort model was run 1,000 times. Mean and standard deviations for the results of the 1,000 replications were calculated. The same outputs were obtained for both multi-cohort and singlecohort model. The multi-cohort model was used to estimate population-level effects, whereas individual benefits and harms were estimated with the same model for a single cohort.

The estimated age-specific BC incidence and mortality during the period 1996–2011, which was a scenario that fit the actual development of the screening programme, was compared with the simulated scenario without screening. Overdiagnosis is defined as the number of women who are diagnosed and treated for cases of breast cancer that never would have become symptomatic in the absence of screening. However, the operational definition in the literature used to estimate overdiagnosis is the difference between the number of BC cases detected with screening and the number without screening. In our case as steady state is not achieved, this definition to estimate overdiagnosis includes not only overdiagnosis but also screen detected BC cases that would be clinically detected in the future in absence of screening (Boer, 1994). Therefore, overdiagnosis is overestimated when analysing population level results in a multi-cohort model that did not achieve the steady state. Accordingly, we will refer in this case to the "incidence increase" instead of "overdiagnosis".

In the case of a single cohort with lifetime follow-up, however the definition used to estimate the incidence increase matches exactly with overdiagnosed cases of BC. We first calculated the relative BC incidence increase (i.e., overdiagnosis) compared with the number of BC cases in a scenario without screening (De Gelder, 2011) and in a second approach, we estimated the fraction of overdiagnosed cases of BC identified by the screening programme.

In addition, the number of women with a false positive result who were referred to the reference hospital for additional tests based on the sensitivity and specificity data from the BCSPBC were also considered harms of the screening programme.

The probabilities of BC-related death in women detected in the screened and unscreened cohorts were analysed with Cox regression only for the single-cohort analysis. Survival time was measured from the beginning of the assigned clinical phase even for BC detected by screening in order to avoid lead time bias.

3.6. Utilities estimation

Due to the lack of quality of live estimations in women affected by BC we decided to apply the methodology described by Stout et al to estimate the age-specific quality-of-life utility weights for the different health states (Stout, 2006). They have fully described the methodology used to estimate age- and stage-specific utilities based on quality of life outcomes reported by general population. Thus the first step consisted in obtaining age-specific EuroQol (EQ-5D) quality-of-life utility weights for general Spanish women population (Oliva-Moreno, 2010).

The study carried out by Oliva-Moreno et al. assessed health related quality of life utilities for general population in the Canary Islands based on the self-perceived health status questions included in the population health survey, specifically EQ-5D questionnaire. The EQ-5 D has five questions asking for a selfperceived status of five different functional conditions related to mobility, personal care, daily activities, pain/discomfort and

anxiety/depression. In each dimension, the interviewed person can choose between three possible answers: 'absence of problems', 'moderate problems' and 'incapacity to perform the activity or severe problems'. A respondent health status is defined by combining one level from each of the 5 dimensions (EQ-5D). A total of 243 possible health statuses can be defined in this way.

In order to translate this number to a single health score, a 'preferences index score or tariff' is needed. Actually, there are two alternative index scores or tariffs validated in Spain, the first one based on a visual analogue scale (the VAS index score or tariff) and the second one based on the time trade-off (TTO index score or tariff (Badia, 2001)). The results derived from both index scores or tariffs are not directly comparable in spite of some attempts to connect them (Dolan, 1997). The results displayed using TTO scale were preferred as final utilities because preferences are usually observed through choices between alternatives health states (Oliva-Moreno, 2010).

Following the mentioned approach, specific percentages were applied to general population utilities in order to estimate the potential negative effects of a BC diagnosis during the first year of treatment and end of life (Table 13). We considered end of life equivalent to the metastatic stage in terms of quality of life and duration.

Table 13: Utilities by breast cancer detection stage and age

Age	Overall population	In situ or Stage II	Stage II or Stage III	Stage IV
50-64	0.8241	0.7417	0.6181	0.4945
65-74	0.7701	0.6931	0.5776	0.4621
75-84	0.6823	0.6140	0.5117	0.4094
>= 85	0.5628	0.5065	0.4221	0.3377
Percentage	100%	90%	75%	60%

We considered that disutilities were only applicable during the first year of treatment after detection, as it is during this first year when women receive the most aggressive treatments. During the following four years in which they continue being treated and have a specific clinical follow-up, up to five years, we assume quality of life decreased to

the level of women treated by breast cancers as in situ (stage 0) or stage I. In addition, even if these women survive more than five years, it is accepted that they will never achieve their quality of life in health state.

On the other hand, in the case of women who relapsed and therefore entered in the phase of terminal care, quality of life would be equivalent to those women in metastatic stage (stage IV). This low level will remain lifetime, both if she finally dies from breast cancer or other causes.

3.7. Costs estimation

Costs related to breast cancer diagnosis were dissagregated in two elements: initial treatment and follow-up. They were estimated based on resources consumption and their unitary costs of the Basque Health Services. Micro-costing technique was applied based on the clinical guidelines used in 2011 in the Basque Health Services (Theriault, 2013). This cost analysis was carried out from the Public Health Service perspective, that is, it considers only those direct costs associated with the delivery of health-related services in relation to the prevention, diagnosis, and treatment of disease. It did not take into account costs related to productivity loss in patients or caregivers, intangible and social costs that are commonly included when social perspective is adopted (Lopez-Bastida, 2010).

Diagnostic costs included medical consultations and confirmation diagnostic tests costs. Mammography costs were obtained from the report published by the BCSPBC (Sarriugarte, 2012). Unitary costs of additional diagnostic test in 2011 were determined using relative value units (RVU) of the Basque Health Service. Specifically ultrasound and fine-needle aspiration cytology (FNAC) costs were calculated based on de RVU assigned for theses procedures on the catalogue of Basque Health Services in 2010. The mean cost for each RVU was 6.31€. Core biopsy cost included the cost of the needle. Finally, open surgical biopsy unitary cost was estimated using the analytical account system together

with the corresponding diagnosis-related group (DRG) generated by the hospitalization (Table 14).

Table 14: Unitary costs and resource consumption for breast cancer diagnosis procedure.

	Resource consumption	Unitary cost	Final cost
Diagnosis			
Medical consultations	3	86	258
Mammogram	1	42	42
Additional diagnosis tests			
Ultrasound	20 %	44	9
Fine-needle aspiration cytology	18 %	113	20
Core biopsy	71 %	127	90
Open surgical biopsy	30 %	2,594	779
Mean diagnosis cost			1,119

Following the methodology described by Theriault et al. (Theriault, 2013), different medical care profiles were defined for stages I to III based on the clinical protocols used in Basque Health Services. Except for the metastasic stage, we distinguished two components in costs: initial treatment and follow-up. Initial costs involved diagnosis, surgery, radiotherapy and chemotherapy while follow-up costs included medical consultations. Metastasic cancers costs (stage IV) were estimated as a unique follow-up category. It was based on a sample of 50 women treated at Cruces Hospital in the Basque Country. Medical consultations, hospitalizations, emergency services, chemotherapy and radiotheraphy resources consumption was recorded during a calendar year in order to estimate the mean annual cost for stage IV.

Those are the main protocol components used for each cancer stage:

- DCIS:
    - Surgery (tumorectomy) – For all.
    - Radiotherapy – For all.

- Stage I:
    - Surgery (tumorectomy or mastectomy with axillary dissection) – For all.
    - Radiotherapy – For all.
    - Chemotherapy – 30% (Negative hormone receptor)
        - FAC: Fluorouracil, Adriamycin, Cyclofosfamide.
        - CMF: Cyclofosfamide, Methotrexate, Fluorouracil.
    - 5 year hormonal treatment – 70%
    - Herceptin (monoclonal antibodies, trastuzumab) – 20% (positive *Human Epidermal Growth Factor Receptor 2* (HER2)).

- Stages II y III:
    - Surgery (tumorectomy or mastectomy with axilar disection) – For all.
    - Radiotherapy – For all.
    - Chemotherapy – 20% (Negative axillary lymph node)
        - FAC: Fluorouracil, Adriamycin, Cyclofosfamide.
    - Chemotherapy – 80% (Positive axillary lymph node)
        - TAC: Docetaxel, Adriamycin, Cyclofosfamide.
        - AC: Adriamycin, Cyclofosfamide + Taxane.
    - 5 year hormonal treatment – 70%
    - Herceptin (monoclonal antibodies, trastuzumab) – 20% (positive *Human Epidermal Growth Factor Receptor 2* (HER2)).

Finally, in stage IV different chemotherapy lines were administrered, mono- or polychemotherapy. In general, almost all treatments include Taxanes and Antracyclines when visceral diseases were present or in the case of negative hormone receptors. In order to calculate surgery related costs tumorectomy and mastectomy costs were estimated with and without axillary dissection. Unitary costs for each surgery intervencion were estimated based on the analytical accounting of the Basque Health Service and weights assigned by the correspondent DRG. Breast reconstruction surgery was not included in these costs.

The unitary cost for each radiotherapy session was available from a report of the Radiotherapy department of the Araba University Hospital. It was based on the analytical accounting system of the hospital (included in the Basque Health Service) taking into account at the same time activity figures, investment made and human resources assigned for year 2011. Investment amortization was hypothesized in 10 years for equipments and in 30 years for the facilities. Radiotherapy session's cost also included outpatient radiotherapy consultations costs and radiophysics support costs.

As chemotherapy costs have high variance, the total cost for each session was estimated summing up pharmacological costs obtained through the Drug Database of the General Council of Pharmacologists (Base de Datos de Medicamentos del Consejo General de Colegios Farmacéuticos, BOT) and the laboratory sale price (LSP) for each cytostatic at January 1st 2011 together with day hospitalization costs and medical consultation carried out before chemotherapy administration. Doses assumed for costs estimation was that for person whose weight is 60 kg and his body surface is 1.6 m^2.

Table 15: Unitary costs and resource consumption related to breast cancer treatment.

	Resource consumption	Unitary cost	Final cost
Surgery			
Tumorectomy	1	3,028	
Tumorectomy + Axillary disection	1	4,072	
Mastectomy + Axillary disection	1	4,072	
	Sessions	Unitary cost	Final cost
Radiotherapy			
Breast protocol	30	227	6.81
Axillary protocol	25	227	5.67
Chemotherapy			
Herceptin	18	1,653	29,754
FAC: fluorouracil, adriamycin, cyclofosfamide	6	244	1,466
FEC: fluorouracil, epirubicin, cyclofosfamide	6	251	1,508
CMF: cyclofosfamide, methotrexate, fluorouracil	6	239	1,436
TAC: docetaxel, adriamycin, cyclofosfamide	6	1,045	6,269

Follow-up treatment costs for stage 0 to stage III were calculated based on hormonotherapy (antiestrogen) use during 5 years and the amount of medical control consultations needed (2 in stage 0 and I; 3 in stage II and 5 in stage III) (Oltra, 2007). Hormonotherapy was recommended for women with positive estrogen receptors, which represent about 70% of treated women. Costs of the antiestrogens that are most commonly used in our Health System (Anastrozole, Letrozole, Exemestane) were obtained from BOT.

Finally, the resulting annual follow-up cost was 1,052€, that taking into account that it was applied only to the 70% of the population, the mean annual cost we determined was 736€.

Table 16: Initial and follow-up costs of breast cancer treatment by detection stage.

	Initial treatment costs				Follow-up costs
Stage	Surgery	Radiotherapy	Chemotherapy	Total	Total
In Situ	3,028	6,810	0	9,838	172
Stage I	4,072	6,810	6,391	17,273	908
Stage II	4,072	6,810	11,263	22,145	994
Stage III	4,072	12,485	12,219	28,776	1,166
	Other annual costs	Annual radiotherapy	Annual chemotherapy	Annual hospitalizations	Annual follow-up costs
Stage IV	790	1,673	13,213	2,203	17,879

Chemotherapy, especially Herceptin (29,754 €), involved the highest costs among the different treatment components displayed in Table 15. However, radiotherapy (6,810 €) also supposed high costs as different surgery types did (Table 16). Initial treatment costs by stage varied from 9,838 € in stage 0 to 24,910 in stage III while follow-up costs ranged from 172€ to 1,116€ for the same detection stages. Chemotherapy (13,213 €) was also the most costly component for women in stage IV. Annual treatment cost for these women was 17,879 €. In the model we calculated the cost according to the duration of stage IV for each woman. As an approximation, in the case of a woman detected in stage

IV whose survival time was 2.8 years (median survival time) yield a cost of 50,061€ patient (Pierga, 2014).

3.8. Cost-effectiveness analysis

Two identical populations were created and followed until death to estimate lifetime costs and QALYs in the screened and unscreened populations. Women in the screened arm were invited according to BCSPBC implementation and no screening mammography was simulated from year 2011 onwards. However, lifetime horizon was applied to the model to include long-term screening effects. According to the approach applied by Stout et al, during this 15-year period (retrospective time), neither costs nor QALYs were discounted, and a 3% annual discount rate was applied prospectively to both costs and QALYs, beginning from the end of the evaluated period (31st December 2011) until death (Stout, 2006; NICE, 2013). In addition, a complementary scenario with no discount (0% discount) applied, neither since 2012, was also considered.

The same model was employed to calculate the ICER for the case of a single cohort of 50,000 women aged 50 years invited to join the programme for the first time in 1996. We used the same alternatives as in the population level approach (with and without screening). As cost-effectiveness analysis is generally applied for a single cohort, these complementary results permit comparison with published data.

3.9. Probabilistic sensitivity analysis

The probabilistic feature of the model was based on varying the main variables randomly at the same time (Briggs, 2006). Each variable was assigned a distribution fitting the range of all possible values and at the beginning of each simulation a random generator selected the value for each variable from the specified distribution. This permitted to

examine the effect of joint uncertainty in the variables of the model through cost-effectiveness plane and acceptability curve (Briggs, 2006). The distributions used for the main parameters varid in the probabilistic sensitivity analysis were detailed in Table 17:

Table 17: Parameters uncertainty included in the probabilistic sensitivity analysis..

Variable	Mean	Distribution
Time between invitations	2.18	Uniform (2.00, 2.36)
Sensitivity	Table 11	Beta (5261, 850)
Specificity	Table 11	Beta (1210790, 100650)
Preclinical state duration	3.44	Uniform (2.88, 4.00)
Screen-detected cancer stage	Table 10	Dirichlet (Table 10)

Time between invitations was calibrated with the aim of reproducing the number of invitations carried out in the programme and the optimal value obtained was 2.18 years. Therefore a uniform distribution was used for this parameter centred in 2.18 and including the theoretical value 2.00 years. The same occurred for the mean value of the duration of the pre-clinical state, where a uniform distribution centred in 3.44, calibrated value, and including 4.00, theoretical value, was used.

On the other hand, a Beta distribution was used both for sensitivity and specificity values. In this case the parameters were based on the number of cases observed in the screening programme in the period 1996-2011: true positive and false negative results for sensitivity and true negative and false positive results for specificity.

Finally, Dirichlet distribution was used for the distribution of detection-stage on screen-detected cancers. The parameters used for Dirichlet were mainly the number of cases observed in the screening programme for each detection-stage depending on the period and detection-age.

The cost-effectiveness plane displays the incremental cost (vertical axis) and effectiveness (horizontal axis) results of 1,000 simulation runs. The mean value and 95% confidence intervals (CI) were shown for the total costs and QALYs, for the differences between the

results for the two scenarios, and for the ICER. In addition, the acceptability curve represents the probability that breast cancer screening is cost-effective compared with no screening for varying threshold values of the cost-effectiveness ratio (Briggs, 2006). The ICER obtained in each of the 1,000 runs is confronted with the different thresholds to calculate those probabilities.

Variability in participation rates was not included in the main probabilistic sensitivity analysis as variability was assumed very small. However, as we were concerned about the interest on the variation of this parameter we ran cost-effectiveness analysis for the main single-cohort model in two more scenarios with lower participation rates: 50% and 30%.

3.10. Budget impact analysis

The simulation model built for multi-cohort cost-effectiveness analysis was used simultaneously for budget impact analysis. Cost-effectiveness analysis allow estimating the additional benefit of a new treatment in relationship with its cost and permit comparing the results to those obtained for already accepted treatments. Undoubtedly, the framework described for cost-effectiveness analysis is accepted by experts panels all over the world (Gold, 1996; Lopez-Bastida 2010). However there are some doubts about its real application when health services management is based on a fixed budget. Budget impact analysis provides a new tool to estimate the effect of the decision hold on the future budget of the health services. As defined by Mauskopf et al. budget impact analysis assesses the impact of a new intervention in annual costs, annual health benefits and other important outcomes from its implementation onwards (Mauskopf, 2005; Sullivan, 2014).

The model was developed to calculate the annual costs for BC diagnosis and treatment in both the screened and unscreened populations. Diagnostic resources included screening or symptomatic mammograms, as well as other additional diagnostic tests that were implemented in the reference hospital. Treatment costs involved the initial treatment of

the BC detected each year and follow-up therapy for prevalent BC, as well as end-of-life costs for those who died from BC. As the budget impact analysis presented financial streams over time, it was not necessary to discount the costs (Sullivan, 2014).

3.11. Individual risk assessment through prediction models

This is a validation study of the main models developed for estimating the risk of BC for women not at high risk. The selected models were identified from the published literature. We included 13,760 women that participated for the first time in the BC early-detection programme in the Sabadell-Cerdanyola (EDBC-SC) area in Catalonia (Spain), between October 1995 and June 1998. The participants did not have a personal history of BC and were followed for vital status or possible diagnosis of BC until July of 2010 (Baré, 2003; Baré, 2006). The EDBC-SC screening programme offers biennial mammography for women aged 50 to 69. The data for this study were obtained through a questionnaire administered on the first visit, which included demographic variables, weight and height, personal gynecological history and family history of BC.

Moreover, as a remarkable and unique characteristic among the Spanish BC screening programs, breast density was recorded on each mammographic test and rated according to the Breast Imaging Reporting and Data System (BI-RADS) (American College of Radiology, 2003). Of the 13,760 women interviewed, we excluded seven without follow-up data, as well as 29 women who were diagnosed with BC and 15 who died within 6 months of baseline. We analyzed incident invasive cancers diagnosed at any time during follow-up, whether the diagnosis was made within the programme or took place outside of it (Baré, 2006). The final sample included 13,709 women, with 329 diagnosed with invasive BC.

Description and changes on the selected models

The models selected for evaluation were developed by Gail (Gail, 1989), Chen (Chen, 2006) and Barlow (Barlow, 2006). The Gail and Chen models have an identical structure.

They estimate the risk of developing BC over time using three components: 1) age-specific relative risks for selected risk factors, 2) incidence of BC in a baseline study population, and 3) competing risks of death.

The original Gail model included both ductal carcinoma in situ (DCIS) and invasive BC. A few years later the incidence rates were modified with the objective of using the model for invasive BC only (Constantino, 1999; Anderson, 1992). Chen and Barlow considered only invasive BC in their respective models. Since the selected models, except the initial Gail model, were developed to predict the risk of invasive BC, in this study we have considered only invasive BC. We customized the Gail and Chen models using an estimated incidence function of invasive BC in Catalonia. Women diagnosed with DCIS in our study cohort were not excluded from the analysis, they were considered at risk of developing invasive BC.

To obtain the baseline BC risk of the study population, required for the Gail and Chen models, BC incidence was multiplied by the complement of the attributable risk (1-AR) corresponding to the distribution of risk factors in the study sample. The AR calculation was performed as described in Chen et al. (Chen, 2006). We used the relative risks of the covariates that were estimated when the models were developed. Since the AR varied little with age, it was considered a constant value for the whole range of ages. The estimated AR for the Gail model was 0.369, and for the Chen model, 0.805. The difference in AR between the Gail and Chen models was due to the fact that the Chen model includes breast density and therefore the baseline risk is considerably lower.

Incidence data for invasive BC were obtained from the Girona and Tarragona Cancer Registries. Incidence rates for the observed period and projected rates for subsequent years were estimated using an age-cohort model with age as a fourth degree and cohort as quadratic polynomials. Breast cancer incidence risk was adjusted by lead time bias and estimated assuming 100% participation rate among 50 years old women invited to the screening programme.

Table 18: Breast cancer incidence rates in Catalonia.

		Birth cohort		
		Year 1930	Year 1940	Year 1950
Age	45-49	93.7	120.6	145.6
	50-54	121.8	156.8	189.3
	55-59	150.5	193.8	234.0
	60-64	184.3	237.4	286.5
	65-69	227.4	292.8	353.4
	70-74	280.3	360.9	435.6
	75-79	334.5	430.7	519.9

Mortality rates in the study (Table 19) population were obtained from the Mortality Registry of the Catalan Department of Health. The mortality rates from causes other than BC, by age and cohort, were obtained from Vilaprinyó et al. (Vilaprinyo, 2008).

Table 19: Other cause mortality rates in Catalonia.

		Birth cohort		
		Year 1930	Year 1940	Year 1950
Age	45-49	229.5	157.7	105.1
	50-54	286.8	196.9	131.1
	55-59	385.4	266.4	178.6
	60-64	550.5	383.8	259.8
	65-69	828.0	532.2	397.6
	70-74	1307.2	925.1	636.2
	75-79	2164.4	1539.1	1063.6

To estimate the relative risks of BC, the Gail model takes into account the number of first degree relatives with a history of BC, age at first live birth, age at menarche and the number of previous benign biopsies. The Chen model also includes breast density and weight, but unlike the initial Gail model does not include the age at menarche or interactions. The Barlow model includes breast density, hormone replacement therapy, body mass index, result of previous mammography exams, race and ethnicity as risk factors. For the Barlow model, the projected risk of BC was based on two separate logistic regression models, one for pre-menopausal women and the other for postmenopausal women.

Projections of risk were obtained at 3 and 5 years, starting six months after the first screening mammogram. Although most of the studies in the literature have worked with five years of follow-up, we considered that projection at 3-years would be useful for short-term decision-making on screening. For the Barlow model, which was designed to estimate the risk of developing invasive BC in a period of one year, the original article recommends projecting the risk for longer periods assuming that the probability of developing BC is identical and independent in each of the ensuing years (Barlow, 2006). Risk estimates for the three models were obtained using Mathematica (Wolfram Research, 2008).

Statistical analysis

We performed a descriptive analysis of the studied variables. Characteristics of women in relation to BC diagnosis were compared using the chi-square test or the Fisher's exact test for dichotomous variables. The calibration of the models was assessed using the Hosmer-Lemeshow goodness-of-fit C statistic (Hosmer, 2000). The C statistic compares the observed (O) and expected (E) number of BC cases by risk quantiles. The expected number of cases was obtained by adding the probabilities estimated by the models for each woman in the group. First, calibration was assessed by quintiles of risk, for the 3 and 5-year projections. Although deciles are often used, we considered that quintiles were more appropriate, given the small number of cancer cases. Then, for the 5-year projections, calibration was assessed on groups determined by categories of risk factors as breast density or family history of BC. Trends in the E/O ratio by categories of risk were

assessed using the chi-square test for trends in order to search for subgroups in which the models worked the best.

The model's discrimination was assessed using the Harrell C statistic, which measures the proportion of all patient pairs in which the predicted breast cancer probability and the follow-up interval (or time to event if the final event occurs), are ranked equally (Harrell, 1982; Harrell, 1996). This concordance measure is a modification of the area under the receiver operating characteristic curve (ROC) that we also included in order to compare our results with similar studies. For these analyses we used the Stata/SE software (StataCorp, 2009).

4. Results

4.1. Epidemiological assessment of the screening programme

Table 20 presents the main population-level results (multi-cohort model) of this evaluation in terms of BC incidence and mortality rates per 100,000 women. BC incidence rates for the age groups in the screening programme were calibrated; thus, goodness of fit between the modelled incidence rates and observed rates was acceptable based on the chi-square test. Compared with the scenario without screening, by 2011, the incidence for women aged 50 to 54 years increased by almost 50% in the scenario with screening, whereas the incidence in age groups between 55 and 69 years increased more than 10%. For women aged 70 to 75 years, BC incidence decreased by 15% because of screening in younger ages.

Table 20: Evolution of estimated breast cancer incidence rates per 100,000 women.

	2000		2005		2011	
BC incidence	Mean	95% CI	Mean	95% CI	Mean	95% CI
Screened population						
<55	316.9	(260.4, 373.4)	285.6	(239.4, 331.8)	310.2	(266.4, 354.0)
55-59	168.9	(138.9, 198.8)	214.9	(181.1, 248.6)	245.9	(209.9, 281.9)
60-64	220.5	(180.6, 260.2)	216.9	(182.0, 251.7)	276.1	(237.0, 315.1)
65-69	162.6	(129.2, 196.0)	199.7	(161.4, 237.8)	264.7	(227.0, 302.3)
70-74	0	(0.0, 0.0)	201.3	(163.3, 239.2)	251.2	(206.4, 295.9)
75-79	0	(0.0, 0.0)	0	(0.0, 0.0)	300.2	(253.5, 346.8)
>=80	0	(0.0, 0.0)	0	(0.0, 0.0)	94.9	(11.7, 178.1)
Unscreened population						
<55	180.2	(137.7, 222.7)	188	(150.5, 225.3)	206.7	(171, 242.3)
55-59	154.5	(125.9, 183.1)	190.8	(159, 222.5)	216.2	(182.4, 249.8)
60-64	204.7	(166.4, 242.9)	200.1	(166.6, 233.4)	249	(211.9, 285.9)
65-69	190.6	(154.4, 226.6)	246.3	(203.9, 288.6)	238.4	(202.7, 273.9)
70-74	0	(0.0, 0.0)	227.2	(186.8, 267.4)	295.9	(247.4, 344.4)
75-79	0	(0.0, 0.0)	0	(0.0, 0.0)	305.8	(258.6, 352.8)
>=80	0	(0.0, 0.0)	0	(0.0, 0.0)	94.9	(11.7, 178.1)

The accumulated increase in BC detection attributable to the screening programme from 1996 to 2011 was 17.0% (Table 21).

Table 21: Accumulated population level results for the period from 1996 to 2011.

	Screened population		Unscreened population	
	Mean	95% CI	Mean	95% CI
Number of women	411,782	(411,619, 411,945)	411,782	(411,619, 411,945)
Participation rate	77.80%	(77.6%, 77.9%)	0%	(0.0%, 0.0%)
Number of mammograms	1,308,030	(1,304,309, 1,311,750)	0	(0, 0)
False positive results	8,211	(7,876, 8,546)	0	(0, 0)
False positive / women	1.99%	(1.91%, 2.08%)	-	-
False positive / mammography	0.63%	(0.60%, 0.65%)	-	-
Screening-detected BC	5,267	(4,999, 5,535)	0	(0, 0)
Total detected BC	10,021	(9,644, 10,399)	8,567	(8,215, 8,918)
Difference in BC detection	1,454	(1,316, 1,593)	-	-
Increase in BC detection	-	-	17.0%	(15.2%, 18.8%)
BC deaths	1,512	(1,370, 1,655)	1,883	(1,725, 2,041)
Difference in BC deaths	-	-	371	(299, 442)
Difference in BC deaths/ BC deaths	-	-	19.7%	(16.3%, 23.1%)

The extension of the programme to women 69 years of age had a big impact in this figure as, in 2005 alone, before the programme was extended, the increase in diagnosed BC cases was 6.3%, whereas using data from 2011, the increase in BC incidence rose to 10.2% (Figure 25).

Among 1,308,030 mammograms performed in the BCSPBC during the evaluated period, 13,478 women with positive mammograms were referred for additional testing at the reference hospital during the study period; 39.1% of them were diagnosed with BC (Table 21).

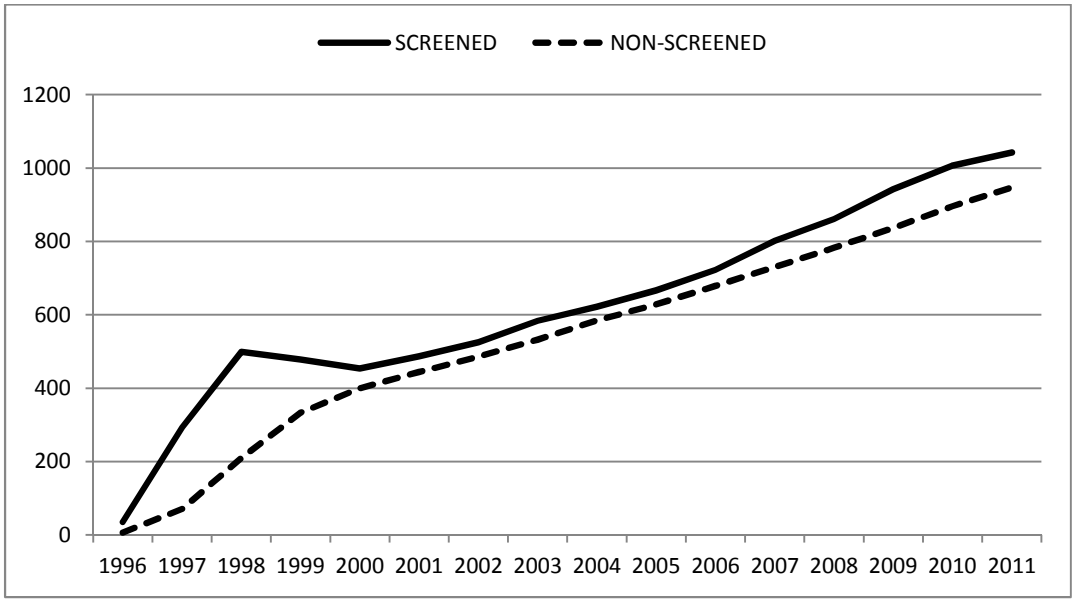


Figure 25: Impact of the screening programme on breast cancer incidence at the population level.

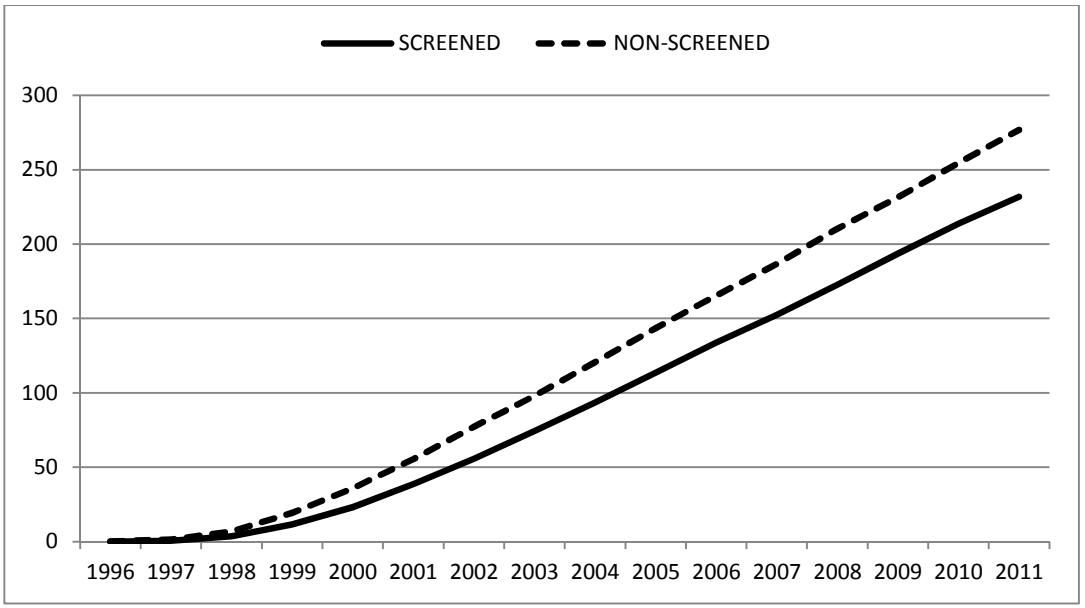


Figure 26: Impact of the screening programme on breast cancer mortality at the population level.

Reduction in mortality was greater the more cohorts were included in the programme. During the study period (1996-2011), the screening programme avoided 19.7% of the deaths due to BC that would have occurred in a scenario without screening (Table 21). Specifically in 2011, the BC mortality rate in women invited into the screening programme was 16.0% lower than in a scenario without screening (Figure 26).

The screening effect on BC mortality in 2011 was greater for women aged 50 to 55 years, with a 27.3% decrease than for the rest of the age groups in the screening programme (55-69 years) with a 22.2% decrease. Women aged 70 or more who previously participated in the programme still had a significant decrease in their probability of death from BC (17.5% for women aged 70-74 years and 2.8% for women aged 75-79), compared with the scenario without screening (Table 22).

Table 22: Evolution of estimated breast cancer mortality rates per 100,000 women.

	2000		2005		2011	
BC mortality	Mean	95% CI	Mean	95% CI	Mean	95% CI
Screened population						
<55	5.2	NA	7.8	(0.1, 15.3)	11.3	(2.9, 19.5)
55-59	8.3	(1.6, 14.9)	27.6	(15.4, 39.6)	35.7	(21.9, 49.4)
60-64	11.2	(2.2, 20.1)	36.4	(22.1, 50.6)	53.2	(36.0, 70.3)
65-69	12.5	(3.2, 21.7)	49.4	(30.4, 68.4)	57.1	(39.6, 74.6)
70-74	0	(0.0, 0.0)	65.2	(43.6, 86.8)	95.5	(67.8, 123.0)
75-79	0	(0.0, 0.0)	0	(0.0, 0.0)	130.3	(99.5, 161.0)
>=80	0	(0.0, 0.0)	0	(0.0, 0.0)	38	NA
Unscreened population						
<55	10.4	(0.2, 20.6)	13.6	(3.5, 23.6)	17.6	(7.2, 28.0)
55-59	12.4	(4.3, 20.5)	37.1	(23.0, 51.0)	45.1	(29.7, 60.5)
60-64	18.6	(7.0, 30.1)	47.8	(31.5, 64.1)	65.8	(46.7, 84.8)
65-69	19.6	(8.0, 31.1)	64.4	(42.7, 86.0)	74.8	(54.8, 94.7)
70-74	0	(0.0, 0.0)	74.5	(51.4, 97.5)	115.9	(85.5, 146.2)
75-79	0	(0.0, 0.0)	0	(0.0, 0.0)	132.1	(101.1, 163.0)
>=80	0	(0.0, 0.0)	0	(0.0, 0.0)	38	NA

When we analysed the survival time, corrected by lead time, for screening-detected BC cases in a single cohort with lifetime follow-up, the hazard ratio for BC mortality was 0.83 (95% confidence interval [CI], 0.77-0.89) (Figure 27).

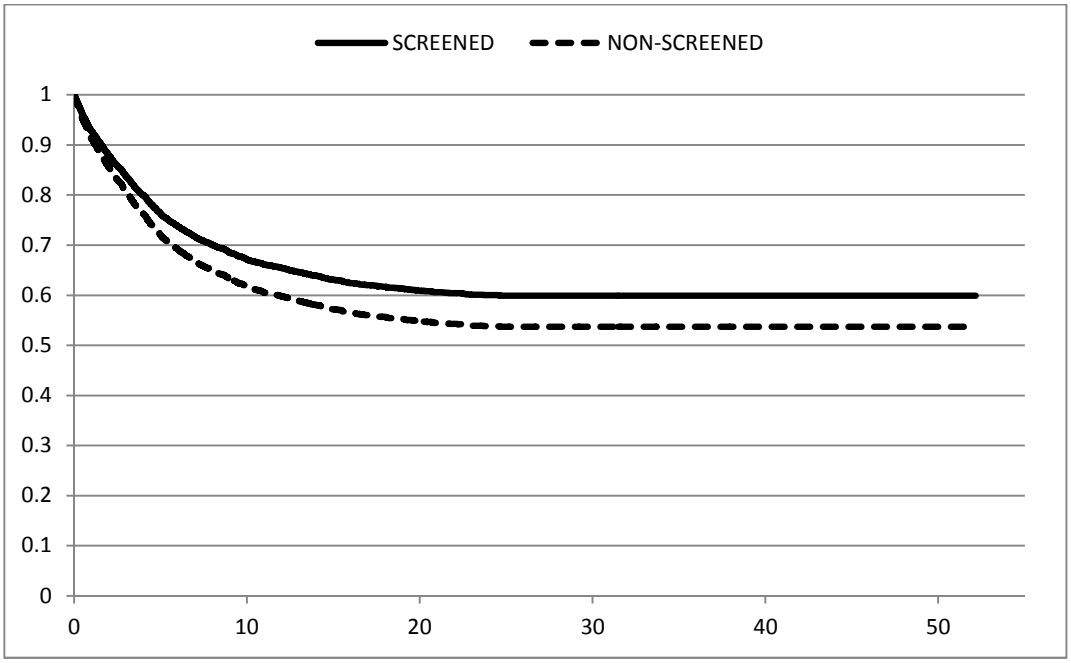


Figure 27: Breast cancer survival analysis corrected by lead time bias.

Life-years gained were 32.6 days for each woman invited into the screening programme and 2.5 years gained for each woman with BC detected by the screening programme (Table 23). In this scenario, 1 out of 28 women detected with BC by screening (3.6%) would be overdiagnosed (Table 23). The mean lead time in BC detection produced by the screening programme, excluding overdiagnosed cases, was 3.2 years (95% CI, 3.1-3.4).

Table 23: Accumulated results for a single cohort of women aged 50 years from 1996 to 2011.

	Screened population		Non-screened population	
	Mean	95% CI	Mean	95% CI
Number of women	50,000	(50,000, 50,000)	50,000	(50,000, 50,000)
Screening age	50-69		-	-
Participation rate	100%		0%	(0.0%, 0.0%)
Number of mammograms	480,869	(480,150, 481,588)	0	(0, 0)
False positive results	3,151	(3,041, 3,260)	0	(0, 0)
False positive / women	6.3%	(6.1%, 6.5%)	-	-
False positive / mammography	0.7%	(0.6%, 0.7%)	-	-
Screen detected BC	1,776	(1,696, 1,856)	0	(0, 0)
Total detected BC	5,065	(4,926, 5,204)	5,001	(4,863, 5,139)
Difference in BC detection (Overdiagnosis)	64	(49, 79)	-	-
Increase in BC incidence	-	-	1.3%	(1.0%, 1.6%)
Overdiagnosis / Screening-detected BC	3.6%	(2.8%, 4.4%)	-	-
BC deaths	1,634	(1,556, 1,711)	1,880	(1,795, 1,964)
Difference in BC deaths	-	-	246	(215, 278)
Difference in BC deaths / BC deaths	-	-	13.1%	(11.6%, 14.6%)
Life years / women	82.6	(82.5, 82.7)	82.5	(82.4, 82.6)
Life years gained / women	0.09	(0.08, 0.10)	-	-
Life years gained / screen detected women	2.5	(2.2, 2.8)	-	-

4.2. Economic assessment of the screening programme

The results of the population-level cost-effectiveness analysis are shown in Table 24. The 15-year evaluation demonstrated a cost of 1,126.6 million euros (1,608.7 million euros, undiscounted) and a provision of 6.70 million QALYs (8.84 million QALYs, undiscounted) for lifetime follow-up. In the non-screened scenario, these values were reduced to 1,090.2 million euros and 6.69 million QALYs. Thus, the incremental cost-effectiveness ratio was 4,214 € per QALY (2,294 €/QALY, undiscounted).

Table 24: Population-level cost-effectiveness analysis for the period from 1996 through 2011.

	0% discount			3% discount		
	Mean	95% CI		95% CI		
Screened population						
Total costs (Million Euros)	1,608.7	1,566.0	1,651.7	1,126.6	1,097.8	1,155.3
Screening mammography costs	55.3	55.2	55.5	55.3	55.2	55.5
Screening diagnosis workup	12.1	11.5	12.7	12.1	11.5	12.7
Clinical cancers diagnosis workup	26.1	25.2	27.0	18.3	17.6	18.9
Treatment costs	1,515.1	1,472.8	1,557.5	1,040.9	1,012.5	1,069.3
QALYs	8,845,493	8,828,791	8,862,195	6,696,959	6,684,899	6,709,019
Unscreened population						
Total costs (Million Euros)	1,584.3	1,538.8	1,629.8	1,090.2	1,059.2	1,121.3
Screening mammography costs	0.00	0.00	0.00	0.0	0.0	0.0
Screening diagnosis workup	0.00	0.00	0.00	0.0	0.0	0.0
Clinical cancers diagnosis workup	30.2	29.2	31.11	22.2	21.5	22.9
Treatment costs	1,554.1	1,509.0	1,599.24	1,068.0	1,037.3	1,098.8
QALYs	8,834,785	8,818,066	8,851,504	6,688,293	6,676,240	6,700,347
Difference (Screened - Unscreened)						
Total costs (Million Euros)	24.4	8.5	40.3	36.4	24.6	1,557.5
Screening mammography costs	55.3	55.2	55.5	55.3	55.2	55.5
Screening diagnosis workup	12.1	11.5	12.7	12.1	11.5	12.7
Clinical cancers diagnosis workup	-4.0	-5.1	-2.9	-3.9	-4.8	-3.1
Treatment costs	-39.0	-54.8	-23.1	-27.1	-38.9	-15.4
QALYs	10,708	9,499	11,917	8,666	7,746	9,586
ICER	2,294	738	3,850	4,214	2,703.41	5,725

When disaggregated costs are analysed, 92% of the total costs were attributed to BC treatment in the screened population. Over the entire study period more than 55 million euros were invested in BC screening mammography, with an additional 12 million for further diagnostic tests, whereas only 4 million euros were saved in clinical or symptomatic diagnosis. Early detection also involved a savings of more than 27 million euros in the treatment of BC detected in the evaluated population.

When a usual single-cohort cost-effectiveness analysis was carried out, the final results were similar in terms of ICER (Table 25).

Table 25: Cost-effectiveness analysis for a single cohort.

	0% discount			3% discount		
	Mean	95% CI		95% CI	95% CI	
Screened population						
Total costs (Million Euros)	213.0	204.7	221.3	161.9	155.9	167.8
Screening mammography costs	12.5	12.458	12.5	12.5	12.5	12.5
Screening diagnosis workup	2.9	2.7	3.1	2.9	2.8	3.1
Clinical cancers diagnosis workup	3.0	2.9	3.2	2.2	2.1	2.3
Treatment costs	194.5	186.3	202.8	144.2	138.3	150.1
QALYs	1,231,858	1,228,748	1,234,968	997,681	995,195	1,000,168
Unscreened population						
Total costs (Million Euros)	206.7	197.4	216.0	153.2	146.5	160.0
Screening mammography costs	0.0	0.0	0.0	0.0	0.0	0.0
Screening diagnosis workup	0.0	0.0	0.0	0.0	0.0	0.0
Clinical cancers diagnosis workup	3.9	3.7	4.1	3.1	2.9	3.2
Treatment costs	202.8	193.6	212.1	150.2	143.5	156.9
QALYs	1,229,578	1,226,441	1,232,715	995,803	993,304	998,301
Difference (Screened - Unscreened)						
Total costs (Million Euros)	6.3	2.5	10.1	8.6	5.7	202.8
Screening mammography costs	12.5	12.5	12.5	12.5	12.5	12.5
Screening diagnosis workup	2.9	2.7	3.1	2.9	2.8	3.1
Clinical cancers diagnosis workup	-0.9	-1.1	-0.7	-0.9	-1.0	-0.7
Treatment costs	-8.3	-12.1	-4.5	-6.0	-8.9	-3.0
QALYs	2,280	1,986	2,575	1,879	1,650	2,108
ICER	2,778	974	4,582	4,623	2,830	6,416

Incremental costs and incremental effectiveness in each of the 1,000 simulations carried out in probabilistic sensitivity analysis are shown graphically in Figure 28. All the simulations resulted in an ICER lower than 10,000€ per QALY.

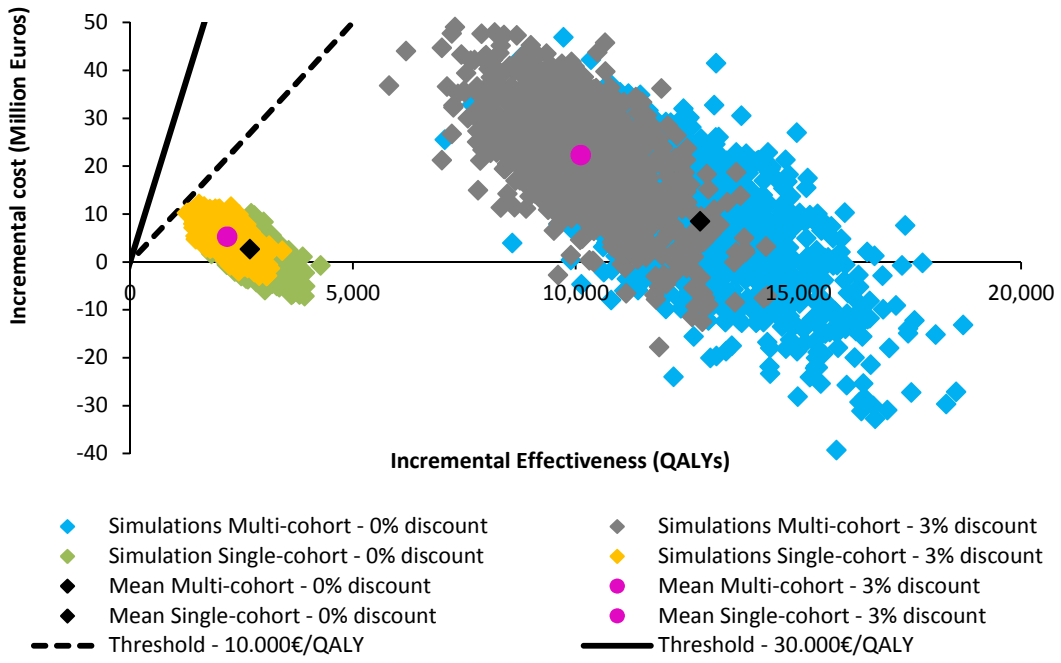


Figure 28: Variability in population-level cost-effectiveness analysis for the period from 1996 through 2011.

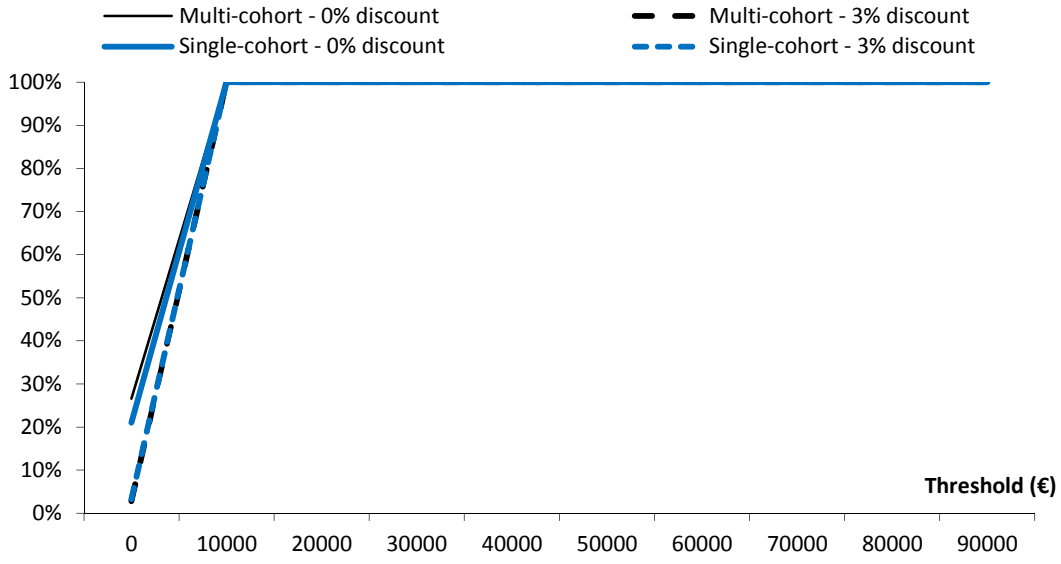


Figure 29: Acceptability curve related to the probabilistic sensitivity analysis.

In addition, the related acceptability curve (Figure 29) showed that in 3% of the simulations screening was dominant (saved costs) both for the single-cohort and multiple-cohort models when no discount was applied. However, this percentage increased up to 21% for the single-cohort model and 27% with population level approach when costs and QALYs were discounted (3% discount).

On the other hand, incremental costs and effectiveness proportionally decreased when lower participation rates were applied in the single-cohort model, therefore the incremental cost-effectiveness ratio result similar in the three scenarios. Main results of the three scenarios were shown in Table 26:

Table 26: Cost-effectiveness analysis for a single cohort in different attendance rate scenarios.

Participation rate	Incremental costs (Million Euros)	Incremental effectivenes (QALYs)	ICER
0% discount			
Base Case	6.3	2,280	2,778
50% attendance	3.2	1,715	1,888
30 % attendance	1.7	1,136	1,453
3% discount*			
Base Case	8.6	1,879	4,623
50% attendance	5.1	1,409	3,601
30 % attendance	2.9	934	3,051

Annual total costs for budget impact analysis are shown in Figure 30. In 2011, more than 36 million euros were necessary to continue with the BCSPBC and the treatment costs related to previously detected BC; this estimation is growing yearly. As a consequence of the implementation of the screening programme, it had been necessary to add up to 9.2 million euros to the budget of the Basque Health Service in 1998. However, this figure became relatively stable from year 2000 onwards in annual 5.2 million euros.

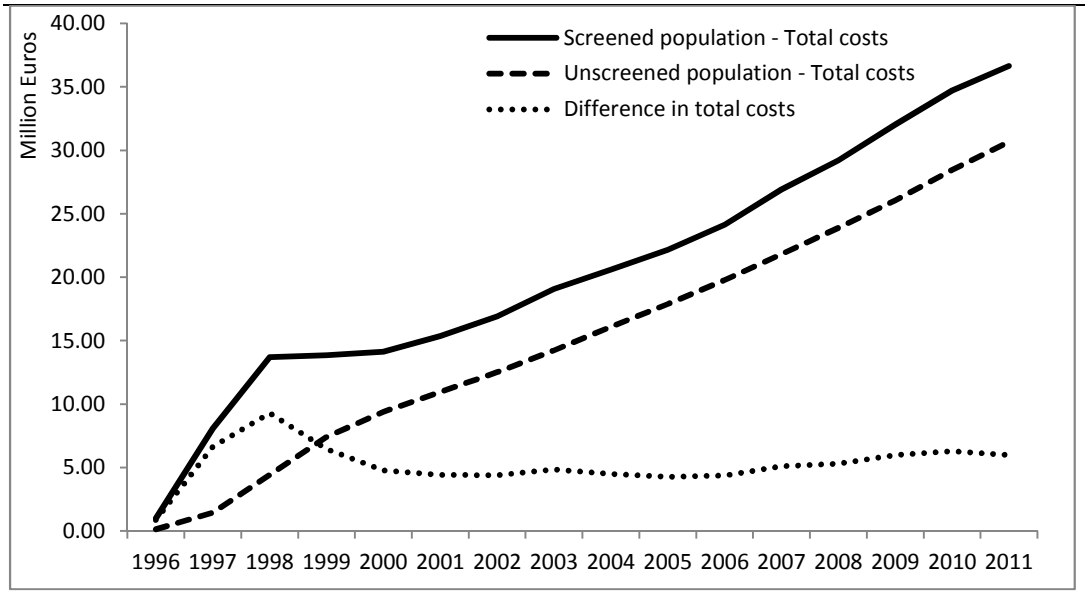


Figure 30: Budget impact analysis for the period from 1996 through 2011.

4.3. Individual risk prediction

Table 27 shows the main characteristics of the studied women. The mean age was 57.0 years and 94.4% of them were postmenopausal. The 18.6% of the women in the study had their first menstrual period before age 12 and the 46.6% of women had their first child at ages between 20 and 24 years. In the study sample, 7.9% of women who subsequently developed invasive BC had first degree relatives with BC while this percentage was 5.3% in women who had not developed BC. This difference was not statistically significant.

However, the differences in breast density, age at first mammogram and previous benign breast disease were significant. Many women reported having previous benign breast disease with no previous biopsy. This was not an unusual practice, in the past, in our publicly funded health system. Median follow-up time was 13.3 years with an interquartile range of 12.7-13.9 years.

Table 27: Characteristics of the study sample used for individual risk models assessment

		No cancer (N=13380)		Cancer (N=329)		Total (N=13709)		
		N	%	N	%	N	%	p-value
Age at first mammogram, y	50-54	4975	37.18	120	36.47	5095	37.17	0.026
	55-59	3602	26.92	110	33.43	3712	27.08	
	60-64	4216	31.51	83	25.23	4299	31.36	
	65-69	587	4.39	16	4.86	603	4.40	
No. of biopsies	0	13263	99.13	325	98.78	13588	99.12	0.513
	>=1	117	0.87	4	1.22	121	0.88	
Menopausal status	No	747	5.58	18	5.47	765	5.58	0.930
	Yes	12633	94.42	311	94.53	12944	94.42	
Age at menarche, y	<12	2493	18.63	62	18.84	2555	18.64	0.990
	12-13	5243	39.19	129	39.21	5372	39.19	
	>=14	5505	41.14	134	40.73	5639	41.13	
	Unknown	139	1.04	4	1.22	143	1.04	
Age at first live birth, y	<20	1065	7.96	26	7.90	1091	7.96	0.050
	20-24	6242	46.65	151	45.90	6393	46.63	
	25-29	4109	30.71	91	27.66	4200	30.64	
	>29	966	7.22	37	11.25	1003	7.32	
	No children	902	6.74	23	6.99	925	6.75	
	Unknown	96	0.72	1	0.30	97	0.71	
Breast density	1	2969	22.19	50	15.20	3019	22.03	<0.001
(BI-RADS) [2]	2	5378	40.19	102	31.00	5480	39.98	
	3	2338	17.47	83	25.23	2421	17.65	
	4	2067	15.45	73	22.19	2140	15.63	
	Unknown	628	4.69	21	6.38	649	4.73	

		No cancer (N=13380)		Cancer (N=329)		Total (N=13709)		
		N	%	N	%	N	%	p-value
Body mass index	<25	3347	25.01	92	27.96	3439	25.09	0.569
	25-29	4519	33.77	105	31.91	4624	33.73	
	30-35	1944	14.53	53	16.11	1997	14.57	
	>35	675	5.04	16	4.86	691	5.04	
	Unknown	2895	21.64	63	19.15	2958	21.58	
Affected first-degree relatives	No	12647	94.52	302	91.79	12949	94.46	0.195
	Yes	704	5.26	26	7.90	730	5.32	
	Unknown	29	0.22	1	0.30	30	0.22	
Previous benign breast disease	No	12420	92.83	292	88.75	12712	92.73	0.005
	Yes	960	7.17	37	11.25	997	7.27	

Table 28A: Calibration of the risk models by quintiles of risk

		N	Expected cases (E)	Observed cases (O)	E/O	C Hosmer-Lemeshow statistic	p-value
Gail model							
3-year	1	2569	10	9	1.11		
	2	2894	13	14	0.93		
	3	2656	13	12	1.08		
	4	2740	16	11	1.45		
	5	2587	20	17	1.18	2.28	0.516
5-year	1	2671	17	16	1.06		
	2	2705	21	22	0.95		
	3	2612	22	16	1.38		
	4	2742	27	18	1.50		
	5	2716	37	35	1.06	4.90	0.180

Validation of the Gail, Chen and Barlow models The Gail and Chen models showed good calibration, at 3- and 5-years, with similar expected and observed number of cases and p-values >0.05 for the Hosmer-Lemeshow C statistics (Table 28). Conversely, the Barlow model overestimated the number of cases, with ratios E/O above 1.8 in all the quintiles of risk and values above 3.3 in the upper quintiles.

Table 28B: Calibration of risk models by quintiles of risk

		N	Expected cases (E)	Observed cases (O)	E/O	C Hosmer-Lemeshow statistic	p-value
Chen model							
3-year	1	2524	7	11	0.64		
	2	2420	9	7	1.29		
	3	2513	12	13	0.92		
	4	2498	15	15	1.00		
	5	2479	22	14	1.57	5.76	0.124
5-year	1	2423	12	14	0.86		
	2	2519	17	11	1.55		
	3	2495	21	21	1.00		
	4	2511	26	23	1.13		
	5	2487	39	28	1.39	5.97	0.113
Barlow model							
3-year	1	2713	19	8	2.38		
	2	2722	29	13	2.23		
	3	2716	39	10	3.90		
	4	2804	50	13	3.85		
	5	2754	70	19	3.68	103.22	< 0.001
5-year	1	2713	31	17	1.82		
	2	2847	51	18	2.83		
	3	2591	62	18	3.44		
	4	2833	84	20	4.20		
	5	2725	115	35	3.29	168.49	< 0.001

When comparing the means of the estimated risk values by BC diagnosis, there were statistically significant differences in the three models at 5-years, but not at 3-years (Table 29). The studied risk models showed poor discrimination in the study sample. The areas under the receiver operating characteristic curve (AUC) ranged from 0.52 to 0.59. For the Gail's model, the AUC confidence intervals for the 3- and 5-year projections included the value 0.50, which indicates the absence of discrimination. The Chen and Barlow models had higher discrimination at five years, with AUCs around 0.58, whereas the Gail model had an AUC around 0.56 in both the 3- and 5-year projections. When time to BC diagnosis was taken into account, the Harrell C statistic indicated that the 5-year projection for the Gail model correctly ordered 56.1% of all pairs of women in the study. The 5-year projection for the Barlow and Chen models increased this figure to 57.5% and 58.6%, respectively (Table 29).

Table 29: Means of the probabilities of developing BC and discriminatory power of the models.

		Gail model	Chen model	Barlow model
3-year	No cancer	0.005340	0.005320	0.015089
	Cancer	0.005728	0.005624	0.015514
	p - value	0.100	0.354	0.648
	AUC	0.562 (0.481, 0.644)	0.523 (0.441, 0.604)	0.526 (0.448, 0.603)
	C-Harrell	0.562 (0.481, 0.643)	0.523 (0.442, 0.603)	0.526 (0.449, 0.603)
5-year	No cancer	0.009226	0.009185	0.024974
	Cancer	0.010014	0.010723	0.028571
	p - value	0.011	< 0.001	0.002
	AUC	0.561 (0.499, 0.623)	0.586 (0.526, 0.646)	0.575 (0.513, 0.638)
	C-Harrell	0.561 (0.480, 0.642)	0.586 (0.526, 0.645)	0.575 (0.513, 0.637)

Table 30 shows the calibration by categories of the risk factors in the studied models. As before, the Gail and Chen models showed good calibration, except for age at first mammogram where the E/O ratio fluctuated. The Barlow model overestimated the number of BC cases and no trends were observed in the categories of the risk factors. By age groups, both the Gail and Chen models overestimated the number of cases in women 50–54 and 60–64 and underestimated them in women 65 years old or older. With regard to breast density, the Gail model overestimated the number of cases in women with densities 1 and 2 and the Chen model in women with breast density 4.

Table 30: Calibration of the risk models by categories of the risk factors.

				Gail's model			Chen model			Barlow model		
		N	Obs.	Exp.	Exp/Obs	p-value	Exp.	Exp/Obs	p-value	Exp.	Exp/Obs	p-value
Age at first mammogram	50-54	5095	31	42	1.35	0.009	41	1.32	0.006	99	3.19	< 0.001
	55-59	3712	34	33	0.97		31	0.91		97	2.85	
	60-64	4299	30	42	1.40		37	1.23		128	4.27	
	65-69	603	13	7	0.54		6	0.46		19	1.46	
Age at menarche	<12	2555	25	25	1.00	0.364	21	0.84	0.552	61	2.44	< 0.001
	12-13	5372	39	50	1.28		45	1.15		134	3.44	
	>=14	5639	43	49	1.14		48	1.12		144	3.35	
Age at first live birth	<20	1091	9	7	0.78	0.181	7	0.78	0.370	24	2.67	< 0.001
	20-24	6393	45	53	1.18		50	1.11		151	3.36	
	25-29	4200	30	42	1.40		38	1.27		104	3.47	
	>29	1003	14	13	0.93		11	0.79		32	2.29	
	No child	925	10	9	0.90		9	0.90		30	3.00	
Body mass index	<25	3439	24	32	1.33	0.200	27	1.13	0.470	85	3.54	< 0.001
	25-30	4624	32	42	1.31		40	1.25		118	3.69	
	30-35	1997	19	18	0.95		19	1.00		53	2.79	
	>=35	691	5	6	1.20		7	1.40		19	3.80	
Breast density (BI-RADS)	1	3019	21	27	1.29	0.063	18	0.86	0.080	37	1.76	< 0.001
	2	5480	36	49	1.36		42	1.17		131	3.64	
	3	2421	29	22	0.76		26	0.90		81	2.79	
	4	2140	17	19	1.12		29	1.71		72	4.24	
No. of biopsies	0	13588	108	123	1.14	-	113	1.05	-	339	3.14	-
	>=1	121	0	1	-		1	-		4	-	
Affected first-degree relatives	No	12949	97	111	1.14	0.149	105	1.08	0.605	317	3.27	< 0.001
	Yes	730	10	13	1.30		10	1.00		23	2.30	
Previous benign breast disease	No	12712	98	115	1.17	0.162	105	1.07	0.754	307	3.13	< 0.001
	Yes	997	10	9	0.90		9	0.90		36	3.60	

5. Discussion

5.1. Main findings

The Breast Cancer Screening Programme in the Basque Country achieved an important reduction in population level BC mortality in 2011 (16.0%), with limited adverse effects. The estimations for 2011 in terms of BC incidence increase and false positive results were in line with the values described in the literature (De Gelder, 2011; Hofvind, 2012; Puliti, 2012; Marmot, 2013; Otten, 2013). In addition, it proved cost-effective during the evaluation period with both multi-cohort and single-cohort approaches. Therefore, our results support the continuation of the programme as a public policy aimed at reducing the burden of BC in the Basque population.

In order to maximize its benefits and minimize harms using existing tools for risk classification would be necessary. However, the three risk predictive models assessed in this study did not show discriminatory power in our setting. Therefore, they cannot be used as a measure of individual risk in early detection programs to customize screening strategies. More work is necessary in this field for obtaining reliable tools to estimate individual risk.

In this line, the BCSPBC has included women aged 40 to 49 years with first-degree family history in the target population since 2012. To be precise, the programme sends a letter to 40 years old women in order to be informed about their family history and lead an early mammography call for those women with affected mother or sisters. Although perfectible, this can be seen as a first step to the presonalization of the screening programme given the lack of adequacy of existing models.

5.2. Epidemiological assessment of the screening programme

Planning a population-level breast cancer screening programme can require multiple criteria for decision making; it involves primarily reduction in BC mortality rates, but it also needs to minimize overdiagnosis and false positive events. Moreover, programme costs must be also taken into account. Decision makers' preferences should be

considered in the interpretation of the results as no unique optimal solution exists (Baltussen, 2006).

Evaluating screening health effects is challenging as we are analysing a system in which the natural history of BC and screening effects are combined. To correctly interpret its impact we should ensure that the measured effects will remain steady in the future; otherwise its evaluation could be misleading. The changes in screening features, such as the extension implemented in 2006, modified the evolution of the indicators (i.e. difference in BC incidence) requires achieving a steady-state to be fully understood. Our results highlight that the variability between the results from different studies depends primarily on the characteristics of the screening programme, but it also depends on the actual variation of these features during its implementation. Therefore, we determined the estimated effects using the model for a single cohort as complementary analysis.

Mortality results are in accordance with the population level results obtained by The Cancer Intervention and Surveillance Modeling Network (CISNET), a consortium of the National Cancer Institute. CISNET results showed a reduction in BC mortality between 7.5% and 22.7% in the United States using seven different models (Cronin, 2006). According to the review published by Broeders et al., the mean European estimate of reduction in BC mortality is 25% to 31% for invited women (Broeders, 2012). Furthermore, a meta-analysis that included 11 randomized controlled trials with 13 years of follow-up estimated a 20% reduction in BC mortality (95% CI 11%-27%) (Marmot, 2013). Specifically, in a study carried out in Spain using mathematical models, Carles et al. concluded a 19.6% reduction in BC mortality, similar to our final estimate (Carles, 2011). When a single cohort (50 years old women invited for the first time in 1996) was analysed with life-time follow up, we concluded that BC mortality decreased by 13.1%. Actually, 2.5 years of life were gained for each woman with breast cancer detected by the screening programme.

The Basque population had not arrived at a steady state by 2011, and therefore, a longer follow-up is necessary to estimate overdiagnosis at the population level without overestimating it. When we assessed 10.2% BC incidence increase in 2011, we included early detected BC cases although they were not necessarily overdiagnoses as they could

be detected from 2012 on. A study published by Duffy et al. (Duffy, 2013) concluded that a 15% increase in incidence in the screened group aged 50 to 69 years can persist 30 years after the start of the programme. In concordance, Yen et al. concluded that there was no excess of incidence in the Swedish Two-County Trial of Mammographic screening at 29 years follow-up and thus, suggest that overdiagnosis is a minor phenomenon (Yen, 2012). Actually, when we analysed a single cohort of 50 years old women followed until death, the estimated increase in incidence was 1.3% of the detected breast cancers compared to a scenario without screening (Boer, 1994). Gunsoy et al. concluded that 5.6% of all BC detected were overdiagnosed in a cohort of British women followed up from age 40 to 85 years (Gunsoy, 2014). The main difference with respect to our 1.3% of overdiagnosed cases might be due to the mean sojourn time in the model of Gunsoy et al. When Gunsoy et al. used a shorter sojourn time, overdiagnosis decreased to 3.1% of all BC diagnosed. Furthermore, as in our model, the cohort was followed until death, the number of all BC detected increased, and therefore, the percentage was lower.

Although overdiagnosis could be considered the major harm of screening, false positive mammographic results are also important harms, since they can cause stress and a decrement in quality of life in healthy women. Our single-cohort analysis reported an accumulated risk of 6.3% for a false positive screening result, with early recall not included as a positive result. False positive rates in our study were concordant with a cumulative chance of 7.3% estimated in the Nijmegen population-based cohort study (Otten, 2013). The literature review by Hofvind et al. (Hofvind, 2012) estimated that the risk of an invasive procedure with benign outcome ranged from 1.8% to 6.3% per woman in Europe.

5.3. Economic assessment of the screening programme

The BC screening programme in the Basque Country proved cost-effective during the evaluation period with both multi-cohort and single-cohort approaches. Although the ICER increased slightly when a 3% discount was applied to costs and utilities from 2011

on, it was far below the recommended threshold of 30,000 € per QALY (Sacristan, 2002). The simultaneous use of a combined and a single-cohort approach was helpful to compare the efficiency of BC screening in real population dynamics (multi-cohort model) and incident cohort (single-cohort). In both cases, the results are valid only if the follow-up is long enough to achieve a steady state in the interaction between the natural history of BC and all its determinants that are modified by the screening. The steady state is defined as the time when each recently observed behaviour of the system (trade-off between short-term costs and long-term benefits) will remain constant in the future (Asmussen, 2007).

In a comparison of different screening programmes, De Koning pointed out the dependence of the cost-effectiveness on the attendance rate and the quality of the programme (De Koning, 2000). Thus, this ICER is within the range of the best programmes as the high participation rate (80%) and other quality indicators of the Basque programme fit well within the range of recommended guidelines (Del Turco, 2010; Canadian Partnership Against Cancer, 2013). As noted in the literature, some of those favourable figures are related to the centralised system applied by the Basque Health Service to implement the BC screening programme (De Koning, 2000). Our results are similar to other studies carried out in the Spanish context that used ordinary, single-cohort cost-effectiveness analysis. Carles et al obtained an incremental cost-effectiveness ratio of 4,469 €/QALY (Carles, 2011) in Catalonia. The MISCAN model applied to Navarra (Van den Akker-van Marle, 1997) resulted in an ICER of 2,650 €/life-year gained (LYG), whereas, when the MISCAN model was applied to Catalonia, it resulted in 4,475 €/LYG (Beemsterboer, 1998). Interestingly, application of the MISCAN model in the Netherlands with the same strategy (women aged 50-70 invited every 2 years) resulted in a similar ICER (3,400 €/QALY) (De Koning, 1991).

Current guidelines for health economic evaluation and modelling have not adequately addressed the issue of cohort definition (Kuntz, 2010). Although the standard approach is to use a single cohort, different authors have underlined the advantages of a multi-cohort method to reproduce real-world populations (Dewilde, 2004; Hoyle, 2010). Our results show that when the same time-horizon and screened women were followed to death, no differences were revealed in the ICER. As Kuntz et al noted, if no substantial

heterogeneity is found on the basis of characteristics of the screened women in the prevalent and incident cohorts, both approaches render similar results (Kuntz, 2010). Similarly, O'Mahony et al. highlighted how the ICER is influenced by the number of birth cohorts under differential discounting (Dewilde, 2004; O'Mahony, 2013). As we have used the same discounting, aggregating cohorts did not produce differences.

All investment decisions involve an opportunity cost, and therefore, a decision to spend on one option deprives the beneficiaries of another option (Gold, 1996). Thus, investment in health care, curative and public health requires evidence of effectiveness and cost-effectiveness of competing interventions (Maynard, 2012). When we take into account both the 67.4 million euros invested in the screening programme during its first 15 years and the total cost of roughly 1,000 million euros, it seems clear that an explicit statement is needed regarding the best use of those resources. Actually, due to the increase in BC incidence and longer survival times achieved by early detection, an increase in the prevalence of treated cancers occurred and thus, overall costs increased considerably. In addition, treatment costs would have continued, even if the screening programme had stopped in 2011. The complementary budget impact analysis showed how the overall annual costs varied in the first years of implementation and the difference between scenarios stabilized after 2000 at approximately 5 million euros. The small increase in 2007 is the result of the increased screening age of 70 years. The overall diagnosis and treatment cost of the BC for the women included in the programme in the Basque Country increased to 36.6 million euros in 2011.

The high attendance rate for the programme helped to reduce disparities in BC survival (Baeten, 2011; Pacelli, 2014). Screening rejection has been proposed on the supposition that new cutting-edge treatments can offset the delay in diagnosis, thus, making it unnecessary to treat at an earlier stage (Biller-Andorno, 2014). This theory has not yet been confirmed, and, even if established, such an approach would not guarantee that innovative therapies would be available to all women with BC. On the contrary, high attendance rates in screening programmes means that the benefit now reaches every female subject in the programme without considering her socioeconomic level.

The retrospective nature of the design of this study posed some doubts about how to deal with discount (Gold, 1996; Stout, 2006; Kuntz, 2010; O'Mahony, 2013). Following the method of Stout et al, we discounted only the future costs and benefits (Stout, 2006). In other words, the results (costs and QALYs) during the evaluation period (1996 to 2011) were directly aggregated, because they had already occurred, but we did discount the follow-up of women living after 2012 to their death as future costs and included QALYs. Although the ICER calculated without any discount changed from 4,214 to 2,294 € per QALY, the difference was not significant, because both figures were far below the usual threshold (30,000 €/QALY). Similarly, from both single-cohort and multi-cohort models, we obtained almost the same ICER (4,600 and 4,200 €/QALY), which underlines the efficiency of the programme.

The growing budget impact indicates that during these years women included in the programme progressively represented a larger portion of the treatment costs of BC. The more years of follow-up included in the programme, the closer the budget is to arriving at a plateau, as these figures include only screened women. These figures highlight the lack of a steady state achieved by the natural history of BC after 15 years of screening.

5.4. Personalizing screening

The principal result of the analysis of predictive models is that when adapting the incidence and mortality rates, the Gail and Chen models were well calibrated to estimate the risk of invasive BC in a population of Spanish women who participated in a screening programme, whereas the Barlow model significantly overestimated this risk. All the three predictive models show a limited level of discrimination, despite the fact that they have been previously used in the US to classify women into high and low risk groups (Constantino, 1999). In general, good performance was seen in the Gail and Chen models when the subgroups of women are defined by categories of risk factors. It is relevant to point out that the use of these models in our study reproduces the original results in terms of discrimination. In the original article, Chen et al. already compared the

discriminatory value of the Gail model against a new model that included breast density. In that case, the AUC for the 5-year prediction was 0.596 for the Gail model and 0.643 for the Chen model (Chen, 2006). In general, it is considered that a prediction tool should have an AUC greater than 0.7 (Hosmer, 2000). With adaptation to the population incidence and mortality rates, we obtained an AUC of 0.561 for the Gail model and 0.586 for the Chen model, for the same 5-year period. Actually, the confidence intervals of the area under the curve in our study contained the values of the original models. The original Barlow publication only showed the discriminatory value of the one-year predictive model, 0.624 (Barlow, 2006). In our study, this figure was 0.602 and the 95% CI (0.440, 0.765) also included the original AUC value.

At the European level, there are adaptations of the Gail model in concrete populations such as an Italian and a Spanish study (Decarli, 2006; Pastor-Barriuso, 2013). One important aspect of these studies is that they include relative risks of the risk factors adapted to their study population. Furthermore, they also modify the incidence of BC as well as mortality by other causes. The risk factors included and the methodology applied for the projection of risks at five years was exactly the same as that used in the original Gail model. Discrimination levels of the Italian and the Spanish adapted models were 0.590 and 0.544, respectively. In the Italian study, the AUC was similar to the 0.586 that Gail found in his study population, whereas in the Spanish study, the AUC was lower and similar to our estimate.

Another article published in the US (Banegas, 2012) showed that the use of relative risks specific to Hispanic and non- Hispanic populations slightly improved discrimination. In our study, the relative risks were not estimated using the study population due to small frequencies in some of the groups defined by risk factors. Although the original relative risks seem to work well for the Gail and Chen models, they may explain in part the lack of calibration of the Barlow model.

Other facts that can explain why the Barlow model did not perform well are differences in the population characteristics, inclusion criteria, and timing projections. In contrast to our study sample, women included in the Barlow study were racially and ethnically diverse. The Barlow study sample included the incident cases detected by the first mammogram

and was developed as a short-time prediction model. Additionally, the model does not use BC incidence rate or mortality by other causes. All these facts also may explain why the Barlow model overestimates risk of breast cancer in our population. A new model for assessing 5-year risk was developed later by the Breast Cancer Surveillance Consortium (Tice, 2008), which would be interesting to assess in a Spanish population in future studies.

In Darabi et al. (Darabi, 2012), where the Gail model was evaluated using data from a Swiss study, the result was an AUC and 95% confidence interval of 0.548 (0.527, 0.568). Furthermore, they determined the improvement in prediction due to the incorporation of breast density and body mass index. The expanded model increased the AUC to 0.571 (0.545, 0.597). Our results show that the Chen and Barlow models, that also incorporate breast density, have slightly greater discriminatory power for prediction at five years than the Gail model. We have identified three published studies in which one of the studied models, the Gail model, was applied to the Spanish population. Pastor-Climente et al. (Pastor-Climente, 2005) estimated the risk of developing BC in a 5-year period, using the Gail model calculator available on the web, without adapting either incidence or mortality for other causes (*Breast cancer Risk Assessment Tool*, National Cancer Institute). The sample used included only women that had been diagnosed with BC. The study concluded that only 42% of women diagnosed with BC had a high risk, defined as 1.67% or greater (Constantino, 1999). Thus, the original Gail model showed low sensitivity, and sensitivity is a required characteristic for a model to be used for decision making in a screening context. Buron et al. (Buron, 2013), in a screening programme context, assessed the utility of the original Gail model to predict BC in women with a prior positive mammogram. At five years, discrimination was low (AUC = 0.61) and, using the standard threshold of 1.67%, sensitivity and specificity were 46.2% and 72.1%, also too low for clinical decision-making. The third study, by Pastor Barriuso et al. (Pastor-Barriuso, 2013), assessed the performance of the original and a recalibrated Gail model together with a new model fully developed by the authors. Consistent with our results, the recalibrated Gail model was well calibrated overall, although it tended to underestimate risk for women in low-risk quintiles and to overestimate it in high-risk quintiles. In our study, we observed concordance between expected and observed in the low-risk groups and a

slight overestimation of risk in high-risk quintiles. Breast density is a risk factor strongly associated with the risk of BC, as demonstrated in recent years in various studies (Chiu, 2010; Boyd, 2011). The Chen model was designed as an adaptation of the Gail model with the incorporation of breast density as a risk factor. If we compare the results obtained in our study, we see that the Chen model shows improved discrimination at five years over the Gail model, although in our sample the Chen model overestimates risk for women with high density. The Chen model used a quantitative measure of density, although it was then categorized into a variable with five categories, similar to the BI-RADS classification. Given the significant correlation between the BI-RADS and other quantitative measurement systems (Martin, 2006; Garrido-Estepa, 2010), and the availability of the BI-RADS in our screening programme, we considered using it as an approximation. Nevertheless, the inclusion of longitudinal measurements of breast density in the models could improve the risk estimates, as other authors have shown (Kerlikowske, 2007). Another risk factor with important weight in these models is family history. The coefficient of the Barlow model, for pre-menopausal women, is similar to the Chen model's coefficient for the variable "number of first-degree relatives with BC". Nevertheless, the Barlow model for post-menopausal women has a lower coefficient. It is possible that part of the risk attributable to family history is explained by other variables, such as body mass index or surgical menopause, which are not included in the other models mentioned. The Gail model, on the other hand, gives a higher weight to family history in comparison to the Chen model. With the inclusion of breast density in the model, family history loses its impact in risk prediction. One of the principal contributions of our study is the assessment of the risk models using specific incidence and mortality rates by birth cohort in our geographic area. This procedure makes it possible to improve the Gail and Chen estimates based on the incidence rates of BC and mortality rates by other causes, which were obtained from a cross-sectional study. Given that BC incidence rates have an increasing trend, cross-sectional rates overestimate rates for past periods and underestimate those of future periods. As a result of using mortality rates by birth cohort, estimated survival in women over 50 in our study increased considerably in comparison with the US data of the original models. Therefore, a conclusion of our study was that, when local data for BC incidence and mortality from other causes were used,

the Gail and Chen models provided unbiased estimates of risk of developing BC in our population.

5.5. Strength and limitations

Our study is strengthened by the exhaustive methodology used for the validation process. The robustness of this study is reinforced by the reproduction of the main results obtained during the study period in the observed screened population (Eddy, 2012). Thus, using the model, we were able to estimate the results in a scenario without a screening programme. The detailed mathematical model built to include the progressive dissemination of screening in all the cohorts between 50 and 69 years of age enabled the interpretation of the mortality reduction achieved in the first 15 years of the programme. Complementary single-cohort analysis was also helpful when interpreting population level effects.

One limitation of this study was that it was built exactly for the Basque women population invited to the BCSPBC from 1996 through 2011. However, using mathematical models for the evaluation of an already implemented screening programme is applicable to other populations.

A second limitation was that long-term follow-up made it necessary to make assumptions for the projected years. In this study the detection stage distribution of symptomatic BC from 2011 on was based on the same parameters as in 2011.

It is worth mentioning also that BC survival functions applied in the simulation model relied on age and stage at detection but did not change during the study years. Consequently, the effect of improved treatment in women with BC was not incorporated. This way we ensure the observed effect was of only related to screening.

On the other hand, in the individual risk models assessment, one limitation was that the population in the BC early-detection programme in the Sabadell-Cerdanyola (EDBC-SC) area used for the analysis was not included in the Girona and Tarragona Cancer

Registries. Although there were no differences observed in incidence rates between Girona and Tarragona, two areas of Catalonia that are geographically separated, it may be that the study area had a lower incidence of BC. Nevertheless, in a previous study, no differences were observed in BC mortality between a geographical region that included the study population, and the provinces of Girona and Tarragona (Perez-Lacasta, 2010).

Other limitations are related to the number of cancer cases and to missing values. As mentioned above, the small number of cancer cases precluded estimating specific relative risks, which have an impact on the performance of the models, along with the incidence and mortality rates. With respect to missing values, a sensitivity analysis with complete data showed that the calibration results were similar and discrimination slightly improved.

Finally, it is worth mentioning that the risk estimates are based only on the baseline characteristics reported at the first screening exam of the early detection programme. With the number of previous biopsies being an important risk factor, a very small number of women reported having had biopsies before their first screening mammography. In these risk models, this is an important issue, because the estimating equation assumes that the probability or the relative risk is maintained over time.

6. Conclusions

This study has assessed the impact of the BCSPBC at the population level in terms of reduction in breast cancer mortality and the number of false positive results and overdiagnosed cases. Fifteen years after the screening programme started, this study supports an important decrease in breast cancer mortality, with reasonable risk of harm with screening. These pidemiological benefits related to the centralised screening system were confirmed by the economic results and sustain the continuation of the breast cancer screening programme in the Basque population.

- The Programme achieved an important reduction in population level BC mortality in 2011 (16.0%), with limited adverse effects. The estimations for 2011 in terms of BC incidence increase and false positive results were in line with the values described in the literature.
- The BCSPBC proved cost-effective during the evaluation period with both multi-cohort and single-cohort approaches. The ICER was far below the recommended threshold of 30,000 € per QALY in all cases.
- The annual cost of BC diagnosis and treatment for the women included in the programme in 2011 was 36.6 million euros approximately 5 million euros more than in the unscreened scenario.
- The three studied risk models do not have discriminatory power in our setting and therefore, they cannot be used as a measure of individual risk in early detection programs to customize screening strategies. More work is necessary in this field for obtaining reliable tools to estimate individual risk.

7. References

American College of Radiology: The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 3rd edition. Reston (VA): American College of Radiology, 2003.

Anderson SJ, Ahnn S, Duff K: NSABP Biostatistical Center Technical Report. Pittsburgh (PA): Department of Biostatistics, University of Pittsburgh; 1992.

Anderson TJ, Lamb J, Alexander F, Lutz W, Chetty U, Forrest AP, et al.Comparative pathology of prevalent and incident cancers detected by breast screening. Edinburgh Breast Screening Project. The Lancet, 1986; 1(8480): 519–523.

Andersson I. Radiographic screening for breast carcinoma. I. Program and primary findings in 45-69 year old women. Acta Radiologica: Diagnosis 1981; 22(2): 185–194.

Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. Journal of the National Cancer Institute Monographs, 1997; 22: 63–67.

Arveux P, Wait S and Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. Eur J Cancer Care, 2003; 12: 143-153.

Ascunce N. Sobrediagnóstico en programas de cribado de cáncer de mama: un efecto adverso inevitable que debe tenerse en cuenta. Med Clin, 2015; 144(4): 161-162.

Asmussen S, Glynn P. Steady-State Simulation. Stochastic Simulation: Algorithms and Analysis. New York.: Springer; 2007.

Azcárate C, Eraso ML, Gafaro A. La investigación operativa en las ciencias de la salud: ¿reconocemos estas técnicas en la literatura actual?. Anales del sistema sanitario de navarra, 2006; 29(3): 387-397.

Badia X, Roset M, Herdman M, Kind P: A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5 D health states. Medical Decision Making, 2001; 21(1): 7-16.

Baeten SA, Baltussen RM, Uyl-de Groot CA, Bridges JF, Niessen LW. Reducing disparities in breast cancer survival--the effect of large-scale screening of the uninsured. Breast J, 2011; 17(5): 548-549.

Baines CJ, Miller AB. The canadian national breast screening study: why it deserves support. Can Fam Physician. 1982 Mar; 28: 481-485.

Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Eff Resour Alloc. 2006; 4: 14–22.

Banegas MP, Gail MH, LaCroix A, Thompson B, Martinez ME, Wactawski-Wende J, et al: Evaluating breast cancer risk projections for Hispanic women. Breast Cancer Res Treat, 2012; 132(1): 347–353.

Baré M, Montes J, Florensa R, Sentís M, Donoso L. Factors related to nonparticipation in a population-based breast cancer screening programme. Eur J Cancer Prev, 2003; 12(6): 487–494.

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. J Med Screen, 2006, 13(4):183–191.

Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst, 2006; 98(17): 1204–1214.

Beaglehole R, Bonita R, Kjellstrom T. Epidemiología básica. Paltex publications. Organización Mundial de la salud, 1994.

Beemsterboer PMM, Warmerdam PG, Boer R, Borras JM, Moreno V, Viladiu P, de Koning HJ. Screening for breast cancer in Catalonia. Which policy is to be preferred? Eur J Public Health, 1998; 8: 241–246.

Biller-Andorno N, Juni MD. Abolishing Mammography Screening Programs? A View from the Swiss Medical Board. N Engl J Med, 2014; 370(21): 1965–1967.

Bjurstam N, Bjorneld L, Duffy SW, Smith TC, Cahlin E, Erikson O, et al.The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49. Journal of the National Cancer Institute. Monographs, 1997; 22: 53–55.

Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, et al. The Gothenburg Breast Screening Trial. Cancer, 2003; 97: 2387–2396.

Boer R, Warmerdam P, De Koning H, Van Oortmarssen G. Extra incidence caused by mammographic screening. Lancet, 1994; 343: 979.

Boer R, Plevritis S, Clarke L. Diversity of model approaches for breast cancer screening: a review of model assumptions by the Cancer Intervention and Surveillance Network (CISNET) Breast Cancer Groups. Statistical methods in medical research, 2004; 13(6), 525-538.

Boyd NF, Martin LJ, Yaffe MJ, Minkin S: Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Res, 2011; 13(6): 223.

Brailsford SC, Harper PR and Sykes J. Incorporating human behavior in simulation models of screening for breast cancer. Eur J Operational Res, 2012; 219: 491-507.

Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Economics, 2006; 15(12): 1295-1310.

Briggs AH, Goeree R, Blackhouse G, O'Brien B. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Med Decis Making. 2002; 22: 290-308.

Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford university press, New York, 2006.

Broeders M, Moss S, Nystrom L, Njor S, Jonsson H, Paap E, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen, 2012; 19 Suppl 1: 14–25.

Buron A, Vernet M, Roman M, Checa MA, Pérez JM, Sala M, et al: Can the Gail model increase the predictive value of a positive mammogram in a European population screening setting? Results from a Spanish cohort. Breast Cancer Res, 2013; 22(1): 83–88.

Canadian Partnership Against Cancer. Report from the Evaluation Indicators Working Group: Guidelines for Monitoring Breast Cancer Screening Programme Performance (3rd Edition). Toronto: Canadian Partnership Against Cancer. 2013.

Carles M, Vilaprinyo E, Cots F, Gregori A, Pla R, Roman R, et al. Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). BMC Cancer, 2011; 11: 192–203.

Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst, 2006; 98(17): 1215–1226.

Chiu SY-H, Duffy S, Yen AM-F, Tabár L, Smith RA, Chen H-H: Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. Cancer Epidemiol Biomarkers Prev, 2010; 19(5): 1219–1228.

Cooper BS. Confronting models with data. J Hosp Infect, 2007; 65 Suppl. 2: 88-92.

Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst, 1999; 91(18): 1541–1548.

Cronin KA, Feuer EJ, Clarke LD, Plevritis SK. Impact of adjuvant therapy and mammography on U.S. mortality from 1975 to 2000: comparison of mortality results from the cisnet breast cancer base case analysis. J Natl Cancer Inst Monogr, 2006; 36: 112–121.

Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K: Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. Breast Cancer Res, 2012; 14(1): R25.

De Gelder R, Heijnskijk EA, Van Ravesteyn NT, Fracheboud J, Draisma G, De Koning H. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev., 2011; 33: 111–121.

De Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JCJM, Collette HJA, Hendriks JHCL, van der Maas PJ. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer, 1991; 49: 531–537.

De Koning HJ. Breast cancer screening; cost-effective in practice? Eur J Radiol, 2000; 33(1): 32-37.

Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH: Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. J Natl Cancer Inst, 2006; 98(23): 1686–1693.

Del Turco MR, Ponti A, Bick U, Biganzoli L, Cserni G, Cutuli B, Decker T, Dietel M, Gentilini O, et al Quality indicators in breast cancer care. Eur J Cancer, 2010 Sep; 46(13): 2344-2356.

Dewilde S, Anderson R. The cost-effectiveness of screening programmes using single and multiple birth cohort simulations: a comparison using a model of cervical cancer. Med Decis Making, 2004; 24: 486–492.

Dolan P, Sutton M: Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. Social Science & Medicine, 1997; 44(19): 1519-1530.

Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. Oxford university press, New York, 2005

Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. Breast Cancer Res, 2013; 15: R41.

Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, 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.

El-Bastawissi AY, White E, Mandelson MT, Taplin SH: Reproductive and hormonal factors associated with mammographic breast density by age (United States). Cancer Causes Control, 2000, 11(10): 955–963.

Ethgen O, Standaert B. Population- versus cohort-based modelling approaches. Pharmacoeconomics, 2012; 30(3): 171-181.

Feuer EJ: Modeling the Impact of Adjuvant Therapy and Screening Mammography on U.S. Breast Cancer Mortality Between 1975 and 2000: Introduction to the Problem. J Natl Cancer Inst Monogr, 2000; 2006(36): 2–6.

Fineberg H. Decision trees: construction, uses, and limits. Bulletin du Cancer, 1980; 67(4): 395-404.

Finkler SA. The distinction between cost and charges. Annals of internal medicine, 1982; 96(1), 102-109.

Forastero C, Zamora LI, Guirado D and Lallena AM. A Monte Carlo tool to simulate breast cancer screening programmes. Physics in Medicine and Biology, 2010; 55: 5213-5229.

Frisell J, Glas U, Hellstrom L, Somell A. Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. Breast Cancer Research and Treatment, 1986; 8(1): 45–54.

Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL: The Wisconsin breast cancer epidemiology simulation model. J Natl Cancer Inst Monogr, 2006; 36: 37–47.

Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst, 1989; 81(24): 1879–1886.

García-Altes A, Navas E, Soriano MJ. Evaluación económica de intervenciones de salud pública. Gaceta Sanitaria, 2011; 25(supl 1): 25-31.

Garrido-Estepa M, Ruiz-Perales F, Miranda J, Ascunze N, González-Roman I, Sánchez-Contador C, et al: Evaluation of mammographic density patterns: reproducibility and concordance among scales. BMC Cancer, 2010; 10: 485.

Gold MR, Siegel JE, Russell LB et al. Cost-effectiveness in health and medicine. Oxford university press, New York, 1996.

Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev, 2013; 6: CD001877.

Gunsoy NB, Garcia-Closas M, Moss SM. Estimating breast cancer mortality reduction and overdiagnosis due to screening for different strategies in the United Kingdom. Br J Cancer, 2014; 110: 2412–2409.

Gutiérrez A., López de Argumedo M, Rico R, Sarriugarte G "Estudio sobre la ampliación de la edad de la población diana del programa de detección del cáncer de mama en la CAPV" Informe de Evaluación. Vitoria-Gasteiz. Departamento de Sanidad, Gobierno Vasco, 2004. Informe nº: Osteba IE-04-02.

Habbema JD, van Oortmarssen GJ, Van der Maas PJ. Mass screening for cancer: the interpretation of findings and the prediction of effects on morbidity and mortality.Clinics in laboratory medicine, 1982; 2(3), 627-638.

Harrell F, Califf R, Pryor D, Lee K, Rosati R. Evaluating the yield of medical tests. JAMA, 1982; 247(18): 2543–2546.

Harrell F, Lee K, Mark D. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med, 1996; 15: 361–387.

Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, et al. False-positive results in mammographic screening for breast cancer in Europe: a literatura review and survey of service screening programmes. J Med Screen, 2012; 19 Suppl 1: 57–66.

Hosmer D, Lemeshow S: Applied logistic regression. New York: Wiley; 2000.

Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Med Decis Making 2010; 30 (4): 426-437.

Izarzugaza MI, Martinez R, Audicana C, Larrañaga N, Hernandez E, Tobalina MC, et al. Cancer in the Basque Country: Incidence, mortality, survival and their trends. Vitoria-Gasteiz: Department of Health and Consumer Affairs, Basque Country Administration; 2010.

Izarzugaza I, Martinez-Cobo R, Cres-Tobalina M, De Castro V, De la Cruz M, Hurtado R, et al. Incidencia del cáncer en la Comunidad Autónoma del País Vasco. 2008-2009.

Vitoria-Gasteiz: Department of Health and Consumer Affairs, Basque Country Administration; 2013

Karnon J, Vanni T. Calibrating models in economic evaluation. Pharmacoeconomics, 2011; 29: 51–62.

Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J, 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.

Kerlikowske K, Ichikawa L, Miglioretti DL, Buist DSM, Vacek PM, Smith-Bindman R, et al: Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. J Natl Cancer Inst, 2007; 99(5): 386–395.

Kuntz KM, Fenwick E, Briggs A. Appropriate cohorts for cost-effectiveness analysis: to mix or not to mix? Med Decis Making, 2010; 30(4): 424-425.

Law AM, Kelton WD. Simulation modeling and analysis. MCGraw-Hill higher education, New York, 2000.

Lee SJ, Zelen M: A stochastic model for predicting the mortality of breast cancer. J Natl Cancer Inst Monogr, 2006; 36: 79–86.

Lopez-Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, Puig-Junoy J. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ. 2010; 11: 513-520.

Mar J, Antoñanzas F, Pradas R, Arrospide A. Modelos de Markov probabilísticos en la evaluación económica de tecnologías sanitarias: una guía práctica. Gaceta sanitaria, 2010; 24: 209-214.

Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer, 2013; 108: 2205–2240.

Marquez-Calderón S, Ladrero-Blasco O, Melendez I, Nuñez-Gallo D, Gonzalez-Aledo A, Cabañas C et al. Documento marco sobre cribado poblacional. Ponencia de cribado poblacional de la comisión de Salud Pública. Ministerio de sanidad y Política social, 2010.

Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, et al: Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. Radiology, 2006; 240: 656–665.

Mauskopf J, Earnshaw S, Mullins CD. Budget impact analysis: review of the state of art. Expert Rew Pharmacoeconomics Outcomes Res. 2005; 5: 65-79.

Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M et al. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices—budget impact analysis. Value in health, 2007; 10(5): 336-347.

Maynard A. Evidence based medicine: an incomplete method for informing treatment choices. Lancet, 1997; 349: 126-128.

Maynard A. Public Health and Economics: A marriage of necessity. Journal of Public Health Research, 2012; 1: e4.

Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. Journal of Medical Screening, 1999; 6(3): 144–148.

National Cancer Institute: Breast Cancer Risk Assessment Tool, SAS codes for Gail model prediction. Available in: http:\\www.cancer.gov\bcrisktool\.

National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. London 2013.

Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet, 2002; 359: 909–919.

Oliva-Moreno J, Lopez-Bastida J, Worbes-Cerezo M, Serrano-Aguilar P. Health related quality of life of Canary Island citizens. BMC Public Health, 2010; 10: 675.

Oltra A, Santaballa A, Munárriz B, Pastor M, Montalar J. Cost-benefit analysis of a follow-up program in patients with breast cancer: a randomized prospective study. Breast J., 2007; 13: 571-574.

O'Mahony JF, van Rosmalen J, Zauber AG, van Ballegoijen M. Multicohort models in cost-effectiveness analysis: why aggregating estimates over multiple cohorts can hide useful information. Med Decis Making, 2013; 33(3): 407-414.

Ormiston-Smith N, Scowcroft H, Thomson CS. Mortality benefits and overdiagnosis estimates for women attending breast screening. Br J Cancer, 2013; 108: 2413–2424.

Otten JD, Fracheboud J, Den Heeten GJ, Otto SJ, Holland R, De Koning HJ, et al. Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study. Ann Oncol, 2013; 24: 2501–2506.

Pacelli B, Carretta E, Spadea T, Caranci N, Di Felice E, Stivanello E, et al. Does breast cancer screening level health inequalities out? A population-based study in an Italian region. Eur J Public Health. 2014; 24(2):280-5.

Parmigiani G, Berry DA, Aguilar O: Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. Am J Hum Genet, 1998; 62(1): 145–158.

Pastor-Barriuso R, Ascunce N, Ederra M, Erdozáin N, Murillo A, Alés-Martínez JE, et al: Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. Breast Cancer Res Treat, 2013; 138(1): 249–259.

Pastor Climente I, Morales Suarez-Varela M, Llopis González A, Magraner Gil JF: Application of the Gail method of calculating risk in the population of Valencia. Clin Transl Oncol, 2005; 7(8): 336–343.

Pérez Lacasta M, Gregori A, Carles M, Gispert R, Martinez-Alonso M, Vilaprinyo E, et al: The evolution of breast cancer mortality and the dissemination of mammography in Catalonia: an analysis by health region. Rev Esp Salud Publica, 2010; 84(6): 691–703.

Perry N, Broeders M, De Wolf C, Tornberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening diagnosis. Health Consumer Protection (ed). Luxemburg: Office for Official Publications of the European Communities; 2006.

Pierga JY, Delva R, Pivot X, Espié M, Dalenc F, Serin D, Veyret C, Lortholary A, Gligorov J, Joly K, Hernandez J, Hardy-Bessard AC. Bevacizumab and taxanes in the first-line treatment of metastatic breast cancer: overall survival and subgroup analyses of the ATHENA study in France. Bull Cancer, 2014; 101: 780-788.

Pinto JL, Sanchez-Martinez FI. Métodos para la evaluación económica de nuevas prestaciones. Centre de recerca en economía i salut, Barcelona. Ministerio de Sanidad y Consumo, 2011.

Puliti D, Duffy SW, Miccinesi G, De Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen, 2012; 19 Suppl 1: 42–56.

Prieto L, Sacristán JA, Antoñanzas F, Rubio-Terres C, Pinto JL, Rovira J. Análisis coste-efectividad en la evaluación económica de intervenciones sanitarias. Med Clín (Barc), 2004; 122(13): 505-510.

Rodríguez-Barrios JM, Serrano D, Monleón T, Caro J. Los modelos de simulación de eventos discretos en la evaluación económica de tecnologías y productos sanitarios. Gaceta Sanitaria, 2008; 22(2): 151-161.

Rojnik K, Naversnik K, Mateovic-Rojnik T, Primiczakelj M. Probabilistic costeffectiveness modeling of different breast cancer screening policies in Slovenia. Value in Health. 2008; 11: 139–148.

Rue M, Vilaprinyo E, Lee S, Martinez-Alonso M, Carles MD, Marcos-Gragera R, et al. Effectiveness of early detection on breast cancer mortality reduction in Catalonia (Spain). BMC Cancer, 2009; 9: 326–335.

Sacristán JA, Oliva J, del Llano J, Prieto L, Pinto JL. ¿Qué es una tecnología sanitaria eficiente en España? Gac Sanit, 2002; 4: 334-343.

Sacristán JA, Ortún V, Rovira J, Prieto L, García-Alonso F. La evaluación económica en medicina. Med Clín (Barc) 2004; 122(10): 379-382.

Sarriugarte G. Generalidades del Programa de Detección Precoz de Cáncer de Mama del País Vasco [Online]. Departamento de Salud del Gobierno Vasco, 2011. Available in:

http://www.osakidetza.euskadi.net/r85-ckenfe11/es/contenidos/informacion/cancer_mama/es_canc_mam/generalidades.html.

Sarriugarte G, Sanz-Guinea A, Mar J, Antoñanzas F, Nuño R, Orue-Etxebarria B, Rueda JR. Estudio de costes del programa de detección precoz del cáncer de mama de la Comunidad Autónoma del País Vasco. Revisión sistemática de estudios de evaluación económica del cribado de cáncer de mama. Investigación Comisionada. Vitoria-Gasteiz. Departamento de Sanidad y Consumo, Gobierno Vasco, 2012. Informe nº: Osteba D-12-03.

Schousboe JT, Kerlikowske K, Loh AJ, 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: 10–20.

Shapiro S, Strax P, Venet L. Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. JAMA, 1966; 195(9): 731–738.

Sociedad Española de Oncología Médica (SEOM). El cáncer en España 2011 [Online]. Available in: http://www.seom.org/en/prensa/el-cancer-en-espanyacom/102744-el-cancer-en-espana-2011.

Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Medical Decision Making, 1993; 13(4): 322-338.

Soto-Alvarez J. Modelos de simulación de eventos discretos: ¿por qué, cómo y cuándo?». Pharmacoeconomics - Spanish Research Articles, 2009; 6(3): 83-89.

Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment. Pharmacoeconomics, 2008; 26(2): 131-148.

StataCorp: Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.

Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst., 2006; 98(11): 774-782

Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, Orlewska E, Penna P, Rodriguez Barrios JM, Shau WY. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health. 2014 Jan-Feb; 17(1): 5-14.

Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, et al.Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. The Lancet, 1985; 1(8433): 829–832.

Tan SYGL, Van Oortmarssen GJ, De Koning HJ, Boer R, Habbema JDF. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr, 2006, 36: 56.

Theriault RL, Carlson RW, Allred C et al; National Comprehensive Cancer Network. Breast cancer, version 3. 2013: featured updates to the NCCN guidelines. J Natl Compr Canc Netw., 2013; 11: 753-760.

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. Ann Intern Med, 2008; 148(5): 337–347.

Titus-Ernstoff L, Tosteson AN, Kasales C, Weiss J, Goodrich M, Hatch EE, et al: Breast cancer risk factors in relation to breast density (United States). Cancer Causes Control, 2006, 17(10): 1281–1290.

Tyrer J, Duffy SW, Cuzick J: A breast cancer prediction model incorporating familial and personal risk factors. Stat Med, 2004; 23(7): 1111–1130.

Van den Akker-van Marle ME, Reep-van den Bergh CM, Boer R, Del Moral A, Ascunce N, de Koning HJ. Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low rate at the second round. Int J Cancer, 1997; 73(4): 464-469.

Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, Legood R. Calibrating models in economic evaluation: A Seven-Step Approach. Pharmacoeconomics, 2011; 29 (1): 35-49.

Vilaprinyo E, Gispert R, Martinez-Alonso M, Carles M, Pla R, Espinas JA, et al. Competing risks to breast cancer mortality in Catalonia. BMC Cancer, 2008; 8: 331–338.

Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. Plos One, 2014; 9(2): e86858.

Weatherly H., Drummond M, Claxton K, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health policy, 2009; 93: 85-92.

Wilson JM, Jungner G. Principles and practice of screening for disease. Public Health Papers 34. Geneva: World Health Organisation; 1968.

Wolfram Research, Inc.: Mathematica version 7. USA: Wolfram Research; 2008.

Yen AM, Duffy SW, Chen TH, Chen LS, Chiu SY, Fann JC, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. Cancer, 2012; 118(23): 5728–5732.

8. Abreviations

AIP: Analisis del impacto presupuestario.

AUC: Area under the curve.

AR: Attributable risk.

BC: Breast cancer.

BCSPBC: Breast cancer screening programme in the Basque Country.

BIA: Budget impact analysis.

BI-RADS: Breast Imaging Reporting and Data System.

CAPV: Comunidad autónoma del País Vasco.

CISNET: The cancer intervention and surveillance modelling network.

CVRS: Calidad de vida relacionada con la salud.

DCIS: Ductal carcinoma in situ.

EDBC-SC: Breast cancer early detection programme in Sabadell-Cerdanyola.

EQ-5D: Euroqol questionnaire with 5 dimentions and 3 levels.

EUSTAT: Basque statistics institute.

ICER: Incremental cost-effectiveness ratio.

GRD: Grupos relacionados por diagnóstico.

NCI: National Cancer Institute.

NICE: National Institute for Clinical Excellence.

PDPCM: Programa de Detección Precoz del Cáncer de Mama.

QALY: Quality adjusted life years.

ROC: Receiver operator characteristic curve.

RR: Relative Risk.

SEOM: Sociedad Española de Oncología Médica.

TAm: Tasa ajustada de mortalidad.

TTO: Time trade-off.

VAS: Visual analogue scale.

9. Acknowledgments

Eskerrak eman batez ere nere gurasoei: Ama, Aita, egindako ahalegina haundia izan da eta bihotzez eskertzen dizuet. Dena unibertsitatean ikasten ez bada ere bidean zerbait ikasiko nuela pentsatu nahi nuke. Amaia eta Mikel ere gertu daude amaieratik, animo!! Eta noski, eskerrik asko zuei ere, aukeran jarrita ere ezingo nuke bidelagun hoberik aukeratu.

Eskerrak baita ere Javierri, Arrasateko ospitalera iritsi nintzen lehenengo egunetik lanbide honetan bidea egin nezan irakasle lanetan aritu delako, pazientziaz, eta eguna joan eguna etorri lan honekiko duen pasioa transmititzen asmatu duelako. Gogoan izan nahi nituzke baita ere Ikerketa Unitateko lankideak eta bidean ezagututako beste hainbat lagun.

Este proyecto fue financiado por el Departamento de Sanidad y Consumo del Gobierno Vasco con el expediente 2010111007.

This study was funded by the grant 2010111007 from the Health Department of the Basque Government.

10.Published papers

- Arrospide A, Rue M, van Ravesteyn NT, Comas M, Larrañaga N, Sarriugarte G, Mar J. Evaluation of health benefits and harms of the breast cancer screening programme in the Basque Country using discrete event simulation. BMC Cancer. 2015 Oct 12; 15: 671.

- Arrospide A, Soto-Gordoa M, Acaiturri T, López-Vivanco G, Abecia LC, Mar J. Cost of breast cancer treatment by clinical stage in the Basque Country, Spain. Rev Esp Salud Pública. 2015 Jan-Feb; 89(1): 93-7.

- Arrospide A, Rue M, van Ravesteyn NT, Comas M, Soto-Gordoa M, Sarriugarte G, Mar J. Economic evaluation of the breast cancer screening programme in the Basque Country: retrospective cost-effectiveness and budget impact analysis. BMC Cancer. 2016. [Under review]

- Arrospide A, Forné C, Rué M, Torà N, Mar J, Baré M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. BMC Cancer. 2013 Dec 10; 13: 587.