

Plataformas genómicas de segunda generación para el pronóstico del cáncer de mama: EndoPredict® y Prosigna™

Revisión sistemática

Second generation prognostic
genomic tests in early breast cancer:
EndoPredict® y Prosigna™
Executive summary

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE SANIDAD, SERVICIOS SOCIALES
E IGUALDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
DE TECNOLOGÍAS Y PRÁCTICAS DE LA SEGURIDAD NACIONAL DE SALUD



JUNTA DE ANDALUCÍA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

Plataformas genómicas de segunda generación para el pronóstico del cáncer de mama: EndoPredict[®] y Prosigna[™]

Revisión sistemática

Second generation prognostic
genomic tests in early breast cancer:
EndoPredict[®] y Prosigna[™].
Executive summary

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA

INFORMES, ESTUDIOS E INVESTIGACIÓN

Martínez-Férez, Isabel María

Plataformas genómicas de segunda generación para el pronóstico del cáncer de mama: EndoPredict® y Prosigna™. Isabel M. Martínez-Férez, Rebeca Isabel-Gómez, Carmen Beltrán-Calvo — Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, 2015.

98 p; 24 cm. (Colección: Informes, estudios e investigación. Ministerio de Sanidad, Servicios Sociales e Igualdad. Serie: Informes de Evaluación de Tecnologías Sanitarias)

ISBN: 978-84-15600-74-9

1. Neoplasias de la mama 2. Perfilación de la Expresión Génica / utilización 3. Genómica / métodos I. Isabel-Gómez, Rebeca II. Beltrán-Calvo, Carmen III. Andalucía. Agencia de Evaluación de Tecnologías Sanitarias IV. España. Ministerio de Sanidad, Servicios Sociales e Igualdad V. España. Ministerio de Economía y Competitividad

Autores: Isabel M. Martínez-Férez, Rebeca Isabel-Gómez y Carmen Beltrán-Calvo

Este documento se ha realizado al amparo del convenio de colaboración suscrito por el Instituto de Salud Carlos III, organismo autónomo del Ministerio de Economía y Competitividad, y la Consejería de Igualdad, Salud y Políticas Sociales de la Junta de Andalucía, en el marco del desarrollo de actividades de la Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del SNS, financiadas por el Ministerio de Sanidad, Servicios Sociales e Igualdad

Edita: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía
Consejería de Igualdad, Salud y Políticas Sociales

JUNTA DE ANDALUCÍA

Avda. de la Innovación s/n, Edificio Arena 1
41020 Sevilla

España – Spain

Teléfono: 955 006 309 Fax: 955 006 327

Mail: aetsa.csbs@juntadeandalucia.es

Web: <http://www.juntadeandalucia.es/salud/aetsa>

ISBN: 978-84-15600-74-9

NIPO: en tramitación

Este documento puede ser reproducido en todo o en parte, por cualquier medio, siempre que se cite explícitamente su procedencia

Plataformas genómicas de segunda generación para el pronóstico del cáncer de mama: EndoPredict[®] y Prosigna[™]

Revisión sistemática

Second generation prognostic
genomic tests in early breast cancer:
EndoPredict[®] y Prosigna[™].
Executive summary

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE SANIDAD, SERVICIOS SOCIALES
E IGUALDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
DE TECNOLOGÍAS Y PRESTACIONES DEL SISTEMA NACIONAL DE SALUD



JUNTA DE ANDALUCÍA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

Contribución de los autores

Isabel M. Martínez-Férez. Doctora en Biología. Agencia de Evaluación de Tecnologías Sanitarias de Andalucía (AETSA). Introducción, metodología, selección de estudios, valoración de la calidad metodológica, extracción de datos y síntesis de los resultados, discusión, conclusiones y revisión del informe final.

Rebeca Isabel-Gómez. Licenciada en Documentación, AETSA. Búsqueda bibliográfica, documentación y revisión del informe final.

Carmen Beltrán-Calvo. Jefa del Servicio de Evaluación de Tecnologías Sanitarias, AETSA. Planteamiento del proyecto, coordinación técnica y revisión del informe final.

Conflicto de Interés

Los autores declaran que no tienen intereses que puedan competir con el interés primario y los objetivos de este informe e influir en su juicio profesional al respecto.

Implicaciones éticas

No se consideraron relevantes los aspectos éticos y legales relacionados con la tecnología evaluada en este informe.

Agradecimientos

Este trabajo se ha beneficiado de forma importante de las aportaciones de los siguientes profesionales:

Dr. Manuel Ruiz Borrego. *Facultativo especialista en Oncología. Unidad de Gestión Clínica Oncología Integral del Hospital. Universitario Virgen del Rocío de Sevilla.*

Dr. Pedro Sánchez Rovira, *Jefe de la Unidad de Gestión Clínica Oncología Radioterápica del Complejo Hospitalario de Jaén.*

Dr. Juan Bayo Calero, *Coordinador de la Sección de Oncología Médica de la Unidad de Gestión Clínica Paciente Oncológico del Hospital Juan Ramón Jiménez de Huelva.*

Soledad Benot López. *Coordinadora de la Unidad de Evaluación de Tecnologías Sanitarias de los Hospitales Virgen Macarena y Virgen del Rocío.*

La Agencia de Evaluación de Tecnologías Sanitarias de Andalucía y los autores agradecen a los revisores de este texto el esfuerzo realizado, su dedicación y sus aportaciones.

Los contenidos del informe son responsabilidad de los autores, procediendo la eximente habitual en el caso de los revisores.

Índice

Índice de tablas y figuras	17
Abreviaturas	19
Glosario	21
Resumen ejecutivo	25
Executive summary	29
1. Introducción	33
1.1 EndoPredict®	35
1.2 Prosigna™	37
Objetivo	40
2. Material y Métodos	41
2.1 Tipo de estudio	41
2.2 Búsqueda bibliográfica: bases de datos y estrategia	41
2.3 Selección de artículos relevantes	41
2.4 Evaluación de la calidad de los estudios	42
2.5 Nivel de evidencia	43
2.6 Extracción y síntesis de los resultados	43
3. Resultados	45
3.1 Resultados EndoPredict®	45
3.1.1 Resultados de la búsqueda	45
3.1.2 Descripción de los estudios y evaluación de la calidad metodológica	46
3.1.3. Resultados clínicos de EndoPredict®	58
3.2 Resultados Prosigna™	64
3.2.1 Resultados de la búsqueda	64
3.2.2 Descripción de los estudios y evaluación de la calidad metodológica	65
3.2.3 Resultados clínicos de Prosigna™	72
4. Discusión	79
5. Conclusiones	85
6. Referencias	87
7. Anexos	93
Anexo 1 Estrategias de búsquedas	93
EndoPredict	93
Prosigna	94
Anexo 2 Listado de comprobación de sesgos de estudios pronósticos ...	95
Checklist: NICE	95

Índice de tablas y figuras

Tabla 1.	Clasificación molecular del cáncer invasivo de mama (Schnitt, 2010)	34
Tabla 2.	Características de las pruebas genómicas EndoPredict® y Prosigna™	39
Tabla 3.	Estudios primarios publicados sobre EndoPredict®	46
Tabla 4.	Características de los estudios de validez clínica de EndoPredict® incluidos en el informe	54-57
Tabla 5.	Resultados del estudio de validación analítica de EndoPredict® de Kronenwett <i>et al.</i> , 2012	58
Tabla 6.	Resultados de validez clínica de los índices EP y EPclin en mujeres post- menopáusicas ER+, HER-2-	60
Tabla 7.	Resultados de validez clínica de los índices EP y EPclin en mujeres con cáncer de mama ER+, HER-2- tratadas con quimioterapia adyuvante a 10 años (Martin <i>et al.</i> , 2014).	61
Tabla 8.	Análisis de subgrupos del estudio de Martin <i>et al.</i> , 2014 en mujeres con cáncer de mama ER+, HER2- tratadas con quimioterapia adyuvante a 10 años	62
Tabla 9.	Comparación de EPclin con los índices clínico-patológicos (Dubsky <i>et al.</i> , 2013b)	63
Tabla 10.	Estudios primarios publicados sobre Prosigna™	65
Tabla 11.	Características de los estudios de validez clínica de Prosigna™ incluidos en el informe	69-71
Tabla 12.	Riesgo de recurrencia a distancia a 10 años en pacientes luminal A y B, con y sin afectación ganglionar	74
Tabla 13.	Supervivencia libre de recurrencia a distancia en los tres grupos de riesgo en la cohorte de ABCSG-8 según el índice ROR	75
Tabla 14.	Supervivencia libre de recurrencia a distancia en los tres grupos de riesgo en pacientes sin afectación y con afectación ganglionar de la cohorte de ABCSG-8 según el índice ROR a 10 años de seguimiento	75
Tabla 15.	Supervivencia libre de recurrencia a distancia por subtipos moleculares en la cohorte de ABCSG-8 según el índice ROR	76
Tabla 16.	Riesgo de recurrencia a distancia por grupos de riesgo en la cohorte mixta TransATAC/ABCSG-8 según el índice ROR a 5 y 10 años de seguimiento	77
Tabla 17.	Comparación de la capacidad pronóstica de varias pruebas genómicas (<i>c-index</i>).	82

Figura 1. Diagrama de la estimación de los índices EP y EPclin
(<http://www.endopredict.com>)..... 36

Figura 2. Diagrama de la tecnología utilizada en la prueba Prosigna™..... 38

Figura 3. Diagrama de flujo de selección de referencias 45

Figura 4. Diagrama de flujo de la selección de estudios de Prosigna™ 64

Abreviaturas

ACB: análisis Coste-Beneficio

ACE: análisis Coste-Efectividad

ACU: análisis Coste-Utilidad

AOL: *Adjuvant! OnLine*

EA: efecto adverso

ECA: ensayo clínico aleatorizado

ER: receptor de estrógeno

70-GS: *70- gene signature* (MammPrint®)

GPC: guía de Práctica Clínica

HER2: receptor 2 de factor de crecimiento epidérmico humano

HR: *hazard Ratio*

HR-positivo ó HR+: receptor hormonal positivo

HR-negativo ó HR-: receptor hormonal negativo

HTA: informes de evaluación de tecnologías sanitarias (*Health Technologies Assessment*)

IR: Índice de recurrencia (en inglés: *Recurrence Score*)

NCCN: *National Comprehensive Cancer Network*

NICE: *National Institute for Health and Care Excellence*

NIHR: *National Institute for Health Research*

NL: nódulo o ganglio linfático

PCR: reacción en cadena de la polimerasa (en inglés: *Polymerase Chain Reaction*)

PR: receptor de progesterona

QT: quimioterapia

RS: revisión sistemática

RT: transcripción reversa (en inglés: *Reverse Transcription*)

RT-PCR: reacción en cadena de la polimerasa y transcriptasa inversa (en inglés: *Reverse transcriptase-PCR*)

TH ó HT: terapia hormonal u Hormonoterapia

TR: tasa de recurrencia

Tto: tratamiento

Glosario

ADN. Abreviatura del ácido desoxirribonucleico (en inglés, **DNA**. *Deoxyribonucleic Acid*). Molécula en doble hélice formada por unidades denominadas nucleótidos que constan de un azúcar (dexosirribosa) fosfatado y una base nitrogenada que puede ser guanina, adenina, timina y citosina. La secuencia de bases nitrogenadas es la que codifica la información genética.

ADNc. ADN complementario formado mediante la transcripción reversa de un ARNm purificado. Este ADN sólo contiene la secuencia codificante del gen del que procede.

Amplificación. Aumento en el número de copias de un fragmento de material genético particular.

ARN. Abreviatura del ácido ribonucleico (en inglés, **RNA**. *Ribonucleic Acid*). Molécula en doble hélice formada por unidades denominadas nucleótidos que constan de un azúcar (ribosa) fosfatado y una base nitrogenada que puede ser guanina, adenina, uracilo y citosina.

ARNm. Abreviatura del ácido ribonucleico mensajero (en inglés, **mRNA**. *Messenger Ribonucleic Acid*). Es la molécula de ARN sintetizada a partir de una secuencia de ADN.

Expresión génica. Formación de un ARNm a partir de un gen.

Factor predictivo. Parámetro que se evalúa para predecir la respuesta probable a un tratamiento específico, por ejemplo, receptor de estrógeno para predecir la respuesta probable a la terapia endocrina.

Factor pronóstico. Características de la enfermedad que se correlacionan con el curso de la enfermedad y que se utilizan para predecir los resultados probables.

Gen. Unidad básica de herencia de los seres vivos.

Hormonoterapia. Tratamiento del cáncer mediante la eliminación y/o el bloqueo de los efectos de las hormonas que estimulan el crecimiento de las células cancerosas. El tratamiento hormonal actúa alterando la producción o impidiendo la acción de los estrógenos o de la testosterona sobre los órganos diana.

Ki67. Marcador de proliferación celular. La proteína Ki67 está presente en todas las fases del ciclo celular excepto en la fase de reposo (G0).

Nódulos o ganglios linfáticos. Glándulas de pequeño tamaño que forman parte del sistema linfático.

Prognosis. Conocimiento anticipado de un suceso. Es la predicción sobre el resultado probable de la evolución de una enfermedad, la posibilidad de recuperación, la recurrencia o la muerte.

Quimioterapia. Tratamiento con fármacos que son tóxicos para las células cancerosas, y que se administra con el objetivo de destruir las células o de prevenir o retrasar su crecimiento.

Reacción en cadena de la polimerasa (PCR). Técnica de biología molecular, cuyo objetivo es obtener un gran número de copias de un fragmento de ADN particular.

Receptor de estrógeno (ER). Molécula (receptor) que se une de forma específica al estrógeno desencadenando cambios dependientes de dicha hormona en el comportamiento de la célula.

Receptor 2 de factor de crecimiento epidérmico humano (HER2). Molécula situada en la superficie celular y que interactúa con el factor de crecimiento específico ayudando a controlar la velocidad a la que crecen las células.

Receptor de progesterona (PgR). Molécula (receptor) que se une de forma específica a la hormona progesterona. El PgR es activado por la hormona progesterona e interviene en la regulación de varias funciones celulares normales, incluyendo la proliferación celular.

RT-PCR. Reacción en cadena de la polimerasa de los productos obtenidos mediante reacción con la transcriptasa inversa. Es una variante de la reacción en cadena de la polimerasa (PCR). La PCR es una técnica de biología molecular que permite sintetizar copias de una secuencia de ADN, utilizando un proceso denominado “amplificación”. En la RT-PCR se parte de una molécula de ARN que mediante una transcripción inversa sintetiza su ADNc usando la enzima transcriptasa inversa, el ADNc resultante se amplifica usando PCR tradicional o en tiempo real.

Transcripción. Proceso por el cual la información genética contenida en una secuencia de ADN se utiliza para sintetizar una molécula de ARNm complementario.

Transcripción inversa. Proceso por el que a partir de una molécula de ARNm se sintetiza la molécula de ADN complementario.

Tratamiento adyuvante. Es el tratamiento que se da además del tratamiento primario (inicial). Está diseñado para ayudar a alcanzar el objetivo principal del tratamiento (la erradicación de la enfermedad). La terapia adyuvante

para el cáncer, por lo general, se refiere a la quimioterapia, la terapia hormonal o radioterapia, cuando se administra después de la cirugía primaria para ayudar a disminuir el riesgo de recurrencia de la enfermedad.

Utilidad clínica. Muestra el balance entre los beneficios y riesgos asociados a la prueba. Indica si la prueba dará lugar a una mejora en los resultados en salud del paciente.

Validez analítica. Es la capacidad de una prueba para medir de manera exacta o precisa el parámetro de interés así como su fiabilidad. En este caso, la medida de interés es la expresión de los ARNm de las células tumorales.

Validez clínica. Es la capacidad de una prueba para identificar de forma precisa y fiable a los pacientes o para predecir los resultados clínicos de interés. Normalmente, se presenta en términos de sensibilidad y especificidad clínica.

Resumen ejecutivo

Título: Plataformas genómicas de segunda generación para el pronóstico del cáncer de mama: EndoPredict® y Prosigna™

Autores: Isabel M. Martínez-Férez, Rebeca Isabel-Gómez, Carmen Beltrán-Calvo.

ANTECEDENTES Y JUSTIFICACIÓN

El cáncer de mama es el cáncer de mayor prevalencia en la población femenina. Con el fin de proporcionar a las pacientes de cáncer de mama una atención más personalizada y adaptada a cada situación, se han desarrollado índices pronósticos que permitan establecer las expectativas sobre la evolución de la enfermedad, y que ayuden a decidir la mejor terapia y atención sanitaria disponible. Las primeras pruebas genómicas pronóstico desarrolladas, Oncotype DX® y MammaPrint®, se basaron en la expresión diferencial, en células sanas y tumorales, de un determinado número de genes. Actualmente, ya están disponibles una segunda generación de pruebas genómicas, entre las que se encuentran EndoPredict® y Prosigna™. Estas pruebas combinan los resultados de expresión génica con parámetros clínico-patológicos, con el objetivo de mejorar la estimación del riesgo de recaída de los pacientes.

OBJETIVO

Valorar la eficacia (en términos de validez analítica y clínica) y la utilidad clínica de las pruebas genómicas EndoPredict® y Prosigna™ como pruebas pronóstico del cáncer de mama en estadios tempranos, frente a los índices pronósticos convencionales de carácter clínico-patológico.

METODOLOGÍA

Se ha llevado a cabo una revisión sistemática de la literatura, para la que se ha realizado una búsqueda bibliográfica en las principales bases de datos biomédicas referenciales de ámbito internacional hasta octubre de 2014. Asimismo, se han consultado las principales Agencias y Unidades de Evaluación de Tecnologías Sanitarias tanto a nivel nacional como internacional. Se ha realizado una lectura crítica de los artículos seleccionados, con el fin de identificar los problemas metodológicos que pudieran influir en su validez, así como una síntesis cualitativa de los resultados.

RESULTADOS

La búsqueda bibliográfica ha permitido identificar 10 referencias de EndoPredict® y 6 referencias de Prosigna™, con información relevante sobre su validez analítica y clínica. Todos los estudios incluidos eran de carácter retrospectivo.

EndoPredict®

Los cuatro estudios localizados sobre validez analítica analizaron la reproducibilidad de la misma. El estudio principal de validación analítica, ha establecido las condiciones de experimentación de la RT-PCR y su reproducibilidad. Se observó una buena correlación entre los valores locales de cada uno de los 7 laboratorios con el de referencia, en un rango coeficiente de correlación de Pearson entre 0,987 y 0,999. Todas las muestras fueron clasificadas correctamente en su grupo de riesgo con una concordancia del 100% y un índice de Kappa de 1.

Los estudios que han valorado la validez clínica de EndoPredict® han mostrado que:

- Los índices EP y EPclin clasificaban los pacientes en grupos de alto y bajo riesgo de recurrencia a distancia, y que existían diferencias significativas entre ellos a 5 y 10 años tanto en mujeres pre y post menopáusicas con o sin afectación ganglionar.
- EndoPredict® parecía clasificar mejor el riesgo de recurrencia que los índices Guía alemana S3 del 2008, NCCN del 2007 y *St Gallen* del 2011.

Solo un estudio ha valorado el impacto de la prueba en la decisión terapéutica, mostrando que la prueba modificaba la decisión terapéutica en el 37,7 % de los pacientes y reducía el número de pacientes tratados con quimioterapia. No se ha localizado evidencia sobre la utilidad clínica de la prueba.

Prosigna™

La información disponible sobre validez analítica de la prueba procede de un único estudio, Nielsen *et al.*, 2014. En cuanto a la precisión de la prueba, se comprobó que tenía una desviación típica inferior a 1 unidad de ROR (escala de 0–100) y que más del 60 % de la variabilidad observada era intrínseca al ensayo. El análisis de fiabilidad de la prueba mostró que el factor de mayor impacto en la varianza de los resultados fue la “sección de tejido” (> 90 % de la varianza). La concordancia, en cuanto al subtipo molecular y al grupo de riesgo, entre los tres laboratorios fue ≥ 90 %. La comparación de

la expresión génica de las réplicas de tejido en cada laboratorio, mostró una alta correlación con un coeficiente de Pearson de 0,98 (IC95 % 0,98—0,98).

La información sobre la validez clínica de la prueba mostró una relación directamente proporcional entre el valor del ROR y el riesgo de recurrencia a distancia, encontrándose una interpretación distinta del índice para pacientes con y sin afectación ganglionar. En los casos con ganglios negativos un riesgo de recurrencia del 10% era pronosticado por un ROR de 42 mientras que este mismo riesgo en pacientes con ganglios positivos es pronosticado por un ROR de 25. El índice ROR ha mostrado que incrementa la información pronóstica aportada por los índices clínico-patológicos a 5 y > 10 años. La clasificación de subtipos establecida por el índice ROR mostró que a 10 y 15 años la supervivencia libre de recurrencia a distancia del subtipo luminal A era mayor que la del subtipo luminal B.

No se ha localizado evidencia sobre el impacto de la prueba en la decisión terapéutica ni sobre su utilidad clínica.

CONCLUSIONES

- Las pruebas genómicas EndoPredict® y Prosigna™ han presentado validez analítica en cuanto a precisión y reproducibilidad entre laboratorios.
- Los índices EP y EPclin de EndoPredict® son factores pronóstico independientes que añaden información pronóstico a los índices clínico-patológicos, en mujeres pre y postmenopáusicas ER-positivo, HER2-negativo con y sin afectación ganglionar, a 10 años de seguimiento. La prueba presenta una buena capacidad discriminatória entre los dos grupos de riesgo que establece: alto y bajo.
- El índice ROR de Prosigna™ es un factor pronóstico independiente que añade información pronóstico a los índices clínico-patológicos, en mujeres postmenopáusicas HR-positivo con y sin afectación ganglionar a 10 años de seguimiento. Muestra una buena capacidad discriminatória entre los tres grupos de riesgo que establece: bajo, intermedio y alto.
- La realización de la prueba EndoPredict® parece tener impacto en la decisión terapéutica, reduciendo el número de pacientes tratados con quimioterapia adyuvante.
- No se dispone de evidencia sobre la utilidad clínica de ninguna de las dos pruebas

Executive summary

Title: Second generation prognostic genomic tests in early breast cancer: EndoPredict® y Prosigna™

Authors: Isabel M. Martínez-Férez, Rebeca Isabel-Gómez, Carmen Beltrán-Calvo.

BACKGROUND AND JUSTIFICATION

Breast cancer is the most prevalent cancer in the female population. In order to provide a more personalized care to breast cancer patients, prognostic indices have been developed to establish expectations about the evolution of the disease and to decide the best therapy and care healthcare available. First prognostic genomic tests, MammaPrint® and Oncotype DX®, were based on the differential gene expression in normal and tumor cells. Currently, a second generation of genomic tests is available, EndoPredict® and Prosigna™, which combine gene expression with clinicopathological parameters in order to improve the estimation of relapse risk of the patients.

OBJECTIVE

To assess the efficacy (analytical and clinical validation) and clinical utility of Prosigna™ and EndoPredict® as prognostic tests in early breast cancer.

METHODOLOGY

A systematic review of the literature was carried out. A literature search was run in the main biomedical data bases until October 2014. In addition, the main national and international agencies and units of Health Technology Assessment were consulted. A critical reading of selected studies was carried out to estimate the quality of their evidence and a qualitative synthesis of the data was performed.

RESULT

Literature search has identified 10 references of EndoPredict® and 6 of Prosigna™ with relevant information about analytical and clinical validity. All studies were retrospective.

EndoPredict®

Four studies on analytical validation have analyzed the test reproducibility and reliability. The main study has provided the RT-PCR performance conditions and its reproducibility in molecular laboratories. Correlation between the reference EP score and the EP scores of seven participating

laboratories was very high, with Pearson coefficients from 0.987 to 0.999. All samples were classified in the right group of risk, showed a concordance of 100% and a Kappa index of 1.0.

Studies about EndoPredict® clinical validation have showed:

- EP and EPclin scores stratified the patients into high-risk and low-risk group of distant recurrence and these risk groups showed statistically significant differences at 5 and 10 year in pre and postmenopausal node-positive or node-negative ER+, HER2-breast cancer patients.
- EndoPredict® scores seemed to improve the patient risk stratification from German S3-2008, NCCN 2007 and St Gallen 2011 guidelines.

Information about impact on clinical decision making is provided for one study. EndoPredict® assay changed the therapy decisions in 37.7% of patients and showed a reduction in the chemotherapy recommendation.

No evidence about clinical utility has been found.

Prosigna™

Analytical validation is provided by Nielsen *et al.*, 2014 study. Precision analysis showed total standard deviation of assay was inferior at 1 ROR unit on a 0–100 scale and more than 60 % of assay variability was due to within-run variance. About reproducibility, the tissue section was the most important source of variation (> 90 % of total variance). For comparison, molecular subtype and risk group, concordance between testing sites was ≥ 90 % and the comparison of gene expression of issue replicates between all sites had a Pearson coefficient of 0.98 (CI95% 0.98–0.98).

Information on clinical validation showed a relationship between the ROR score and the distant recurrence risk in postmenopausal women with HR-positive, node-negative, node-positive early breast cancer at 10 years. Patients in the ROR high-risk group had higher risk of distant recurrence than patients in the ROR intermediate-risk group and similarly, patients in the ROR intermediate-risk group had higher risk of distant recurrence than patients in the ROR low-risk group.

The ROR score showed a different interpretation for node-negative patients and node-positive patients; for negative lymph a recurrence risk of 10% was predicted for a ROR value of 42 while this same risk in patients with positive nodes is predicted for a ROR value of 25.

ROR score added significant prognostic information to clinicopathological factors on distant recurrence at 5 and 10 years. In addition, Luminal A (ROR molecular subtypes) showed a distant recurrence-free survival significant higher than Luminal B, at 10 and 15 years.

No evidence, about impact on therapeutic decision making and clinical utility of assay, has been found.

CONCLUSIONS

- EndoPredict® and Prosigna™ have showed analytical validity, measured as precision and reproducibility of assay.
- EndoPredict® EP and EPclin scores are independent prognosis factors. These scores provided additional prognostic information on distant recurrence to the clinicopathological factors in pre and postmenopausal women with ER-positive, HER2-negative, node-negative early breast cancer at 10 years. The assay showed a good discrimination capacity between the low and high distant recurrence groups.
- Prosigna™ ROR score is an independent prognosis factor and adds prognostic information on distant recurrence to the clinicopathological factors in postmenopausal women with HR-positive, node-negative, node-positive early breast cancer at 10 years. The assay showed a good discrimination capacity between the three distant recurrence groups: low, intermediate and high.
- EndoPredict® seemed to have an impact on therapeutic decision making, reducing chemotherapy.
- The clinical utility of both assays, and so their potential benefits in breast cancer remain unclear.

Introducción

El cáncer de mama es el cáncer de mayor prevalencia en la población femenina. En Estados Unidos, según datos publicados en 2014, el cáncer de mama representaría el 29 % de total de nuevos casos de cáncer diagnosticados en mujeres (Siegel, Ma, Zou & Jemal, 2014). En España, en los últimos datos publicados correspondientes al año 2012 se ha registrado que el cáncer de mama ha sido el cáncer de mayor incidencia, mortalidad y prevalencia a 5 años (29 %, 15,5 % y 40,8 %, respectivamente) en la población femenina (Sociedad Española de Oncología Médica, 2014). La complejidad de la enfermedad y su heterogeneidad dificultan la valoración tanto de su progresión como de la respuesta a los distintos tratamientos disponibles.

La evolución hacia una medicina personalizada supone un reto para la práctica clínica habitual al implicar una visión diferente de la asistencia sanitaria, centrándose en el paciente. Dentro de esta tendencia, las técnicas que faciliten la clasificación o identificación más precisa de cada paciente, y por consiguiente, permitan establecer la terapia más adecuada a sus características concretas, están en pleno desarrollo.

Con el fin de proporcionar a las pacientes de cáncer de mama una atención más personalizada y adaptada a cada situación, se han desarrollado índices pronósticos que permitan establecer las expectativas sobre la evolución de la enfermedad en cada caso, y que por lo tanto, ayuden a decidir la mejor terapia y la atención sanitaria disponible. Los índices pronósticos más utilizados en el cáncer de mama son: *NCCN guidelines*, *St. Gallen Expert Criteria*, *Adjuvant! Online*, *Nottingham Prognostic Index*. Todos ellos están basados en factores de riesgo de carácter clínico-patológicos como: tamaño y grado del tumor, afectación de los nódulos linfáticos, receptores hormonales (ER/PG) y receptores HER2, la edad de la paciente, menopausia, comorbilidades... El valor pronóstico de estos índices es muy limitado, así, tumores con similares características clínico-patológicas pueden evolucionar de manera muy diferente.

El desarrollo de las técnicas de biología molecular ha permitido estudiar las características moleculares de la enfermedad mediante la expresión diferencial de determinados genes, en tejidos sanos y tejidos tumorales. Los análisis moleculares de la expresión genética han abierto una nueva vía de conocimiento en el complicado comportamiento de las células cancerosas y han permitido identificar posibles genes asociados al comportamiento de las células tumorales.

La incorporación de las técnicas moleculares a la práctica clínica ha facilitado el diagnóstico de un amplio número de enfermedades originadas por alteraciones genéticas, y están siendo utilizadas en otros campos además

del diagnóstico, incorporándose como herramientas importantes dentro del campo pronóstico y predictivo. El conocimiento de las características genéticas de cada enfermedad puede ayudar en la identificación de aquellos pacientes susceptibles de responder adecuadamente a determinados tratamientos y de esta manera, poder decidir la terapia más adecuada en cada caso. Por consiguiente, los avances en la genética humana están produciendo un impacto en la atención sanitaria, que tiende a ofrecer una atención más personalizada con actuaciones preventivas y terapéuticas adaptadas al perfil genético de cada individuo.

Para el cáncer de mama en concreto, la incorporación de las técnicas genéticas a la práctica clínica ha puesto de manifiesto la existencia de diferentes subtipos moleculares de cáncer de mama. Así, se han identificado 3 subtipos moleculares de cáncer: luminal, HER2 y basal (Schnitt, 2010).

Tabla 1. Clasificación molecular del cáncer invasivo de mama (Schnitt, 2010)			
	Subtipos moleculares		
	Luminal	HER2	Basal
Perfil de expresión génica	ER (+) y/o PR (+) y otros genes asociados Tipo A: HER2 (-) Tipo B: HER2 (+)	HER2 (+) ER (-) / PR(-)	HER2 (-) / ER (-) / PR(-) CK (+) EGFR (+)
Características Clínicas	70 % de los cánceres invasivos Luminal B tiende a ser de grado histológico mayor que el tipo A	15 % de los cánceres invasivos En muchos casos tumores de grado histológico alto y nódulos afectados	15 % de los cánceres invasivos Se denomina triple negativo Más frecuente en mujeres afroamericanas
Respuesta a tratamiento	Responden a terapia hormonal Respuesta variable a quimioterapia (mayor respuesta en el tipo B) El tipo A tiene mejor pronóstico que el B	Responden a trastuzumab y a quimioterapia basada en antraciclina	No responde a terapia hormonal ni a trastuzumab
Pronóstico	Bueno	Generalmente malo	Generalmente malo

El subtipo luminal caracterizado por la expresión de genes receptores de hormonas (ER+ y/o PR+) se clasifica a su vez en dos, luminal A y luminal B según sean HER2- o HER2+, respectivamente. Esta identificación molecular parece ayudar a la hora de conocer el pronóstico de la enfermedad y la respuesta a tratamiento (Zhang, Man, Zhao, Dong & Ma, 2014).

La clasificación molecular establecida por el consenso internacional de expertos de St. Gallen en 2013, diferencia el subtipo luminal A y B en función del valor de Ki67 considerando que en el subtipo Luminal A los valores de Ki67 son bajos. Aunque la mayoría de los expertos consideraron que un valor de Ki67 de ≥ 20 % indicaba valores altos de Ki67, existe controversia entre los expertos en cual debe de ser el punto de corte que diferencie entre valor bajo y alto de Ki67. Además, se diferenciaron dos tipos de Luminal B, un tipo HER2 positivo y otro tipo HER2 negativo (Goldhirsch *et al.*, 2013).

Las pruebas genómicas de carácter pronóstico y/o predictivo en cáncer diseñadas para estimar el riesgo de recurrencia a distancia de la enfermedad, y/o predecir con la mayor precisión posible la respuesta a la terapia adyuvante, tienen como finalidad, la estratificación o clasificación de los pacientes en diferentes subtipos genéticos que recibirían un tratamiento de manera individualizada y más adecuada a cada uno de ellos.

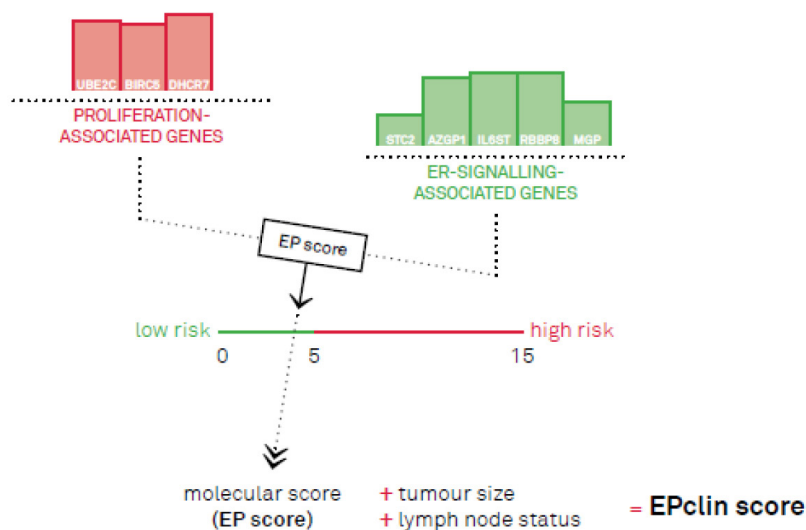
Varias de estas pruebas pronóstico están siendo comercializadas e introducidas en la práctica clínica. Las primeras pruebas desarrolladas, Oncotype DX® y MammaPrint®, se basaron en la expresión diferencial, en células sanas y tumorales, de un determinado número de genes (21 y 70 respectivamente). Actualmente, ya están disponibles y en fase de introducción una segunda generación de pruebas genómicas, entre las que se encuentran EndoPredict® y Prosigna™. Estas pruebas se denominan de segunda generación ya que combinan los resultados de expresión génica con parámetros clínico-patológicos, en un intento de afinar y mejorar la estimación del riesgo de recaída de los pacientes. Una característica adicional de estas pruebas de segunda generación es la descentralización de su realización. En el caso de las pruebas genómicas Oncotype DX® y MammaPrint® las muestras tumorales son analizadas en los laboratorios de referencia propiedad de las compañías que las han desarrollado localizados en Estados Unidos y Holanda, respectivamente. Por el contrario, EndoPredict® y Prosigna™ son pruebas diseñadas para poder ser realizadas en laboratorios locales disminuyendo de esta manera el tiempo de espera del resultado y la posibilidad de realizar la repetición de la prueba en caso de duda. Sin embargo, dada la naturaleza de las pruebas, esta descentralización llevaría implícita la formación específica del personal en los laboratorios, así como el mantenimiento de los equipos.

1.1 EndoPredict®

Es una prueba genómica basada en la expresión de 8 genes, 3 del ciclo de proliferación celular y 5 de señalización hormonal, 1 de estos últimos está relacionado con el síndrome metabólico. Además, la prueba incluye 4 genes control. Se ha diseñado para establecer el riesgo de recurrencia a distancia

de la enfermedad en mujeres, pre y postmenopáusicas, con cáncer de mama ER+, HER2- en estadios tempranos con y sin afectación ganglionar. Esta prueba genómica establece, según los niveles de expresión de los genes antes mencionados, un índice denominado EP que presenta un rango de valores entre 0 y 15. La clasificación del riesgo, en bajo y alto, se realiza considerando el valor 5 como punto de corte entre los dos niveles de riesgo. Este índice se combina con dos variables clínicas, tamaño del tumor y estado de afectación ganglionar, obteniéndose un nuevo índice denominado EPclin. El EPclin aporta una clasificación del riesgo más precisa, incrementando la capacidad pronóstica del EP (Figura 1).

Figura 1. Diagrama de la estimación de los índices EP y EPclin (<http://www.endopredict.com>).



Los análisis de expresión génica se realizan en laboratorios locales de patología molecular; posteriormente, los valores obtenidos en el análisis de expresión son introducidos en la página web generadora de los informes, en la que se incorporan los archivos de texto con los valores de Ct del software VERSANT® y se cumplimentan una serie de campos obligatorios tras lo cual, se genera el informe de la paciente (<http://www2.endopredict.com/EPReportGenerator/>).

La tecnología utilizada en la prueba para determinar la expresión génica es una RT-PCR cuantitativa con sondas taqman a partir de material procedente de la pieza postquirúrgica en parafina.

1.2 Prosigna™

El ensayo Prosigna™ se ha diseñado como un ensayo diagnóstico *in vitro* que analiza el perfil de expresión génica en células cancerosas de tejido de mama, con la finalidad de valorar el riesgo de recurrencia a distancia de los pacientes. Al igual que la prueba EndoPredict®, este ensayo incorpora tanto información de la expresión génica como variables clínicas, y mediante un algoritmo establece el riesgo de recaída o recurrencia de la enfermedad.

El análisis de expresión génica utilizado en la prueba es el PAM50 y comprende la expresión génica de 50 genes, este análisis permite clasificar el tumor de mama en uno de los cuatro subtipos moleculares: luminal A, luminal B, HER-2 y basal.

El algoritmo propuesto por Prosigna™ combina datos genómicos basados en el subtipo molecular del tumor (PAM50) y la información clínica correspondiente a tamaño del tumor, estado de proliferación del tumor y el estado de afectación ganglionar. Este algoritmo proporciona un índice denominado ROR, cuyo valor está comprendido entre 0–100, y clasificando a los pacientes en tres categorías de riesgo: alto, bajo e intermedio. Los puntos de corte que marcan los tres grupos de riesgo difieren en función de la existencia de afectación ganglionar. Por consiguiente, un mismo valor de ROR representa distinto riesgo de recurrencia a distancia a 10 años si el paciente tiene o no afectación ganglionar.

Sin afectación ganglionar:

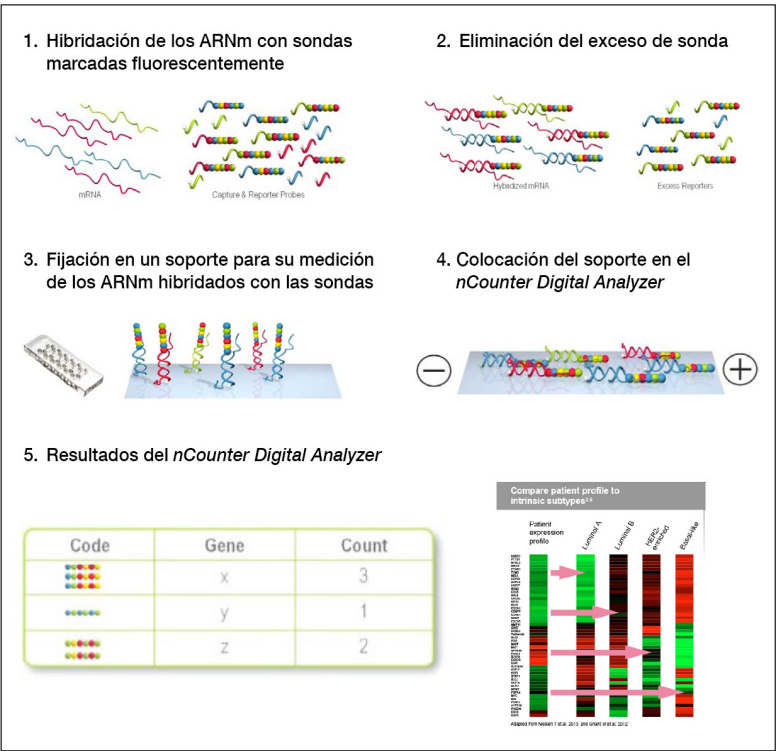
- Riesgo bajo: ROR de 0–40
- Riesgo intermedio: ROR de 40–60
- Riesgo alto: ROR de 60–100

Con afectación ganglionar:

- Riesgo bajo: ROR de 0–15
- Riesgo intermedio: ROR de 15–40
- Riesgo alto: ROR > 40

Este ensayo está indicado para mujeres postmenopáusicas con cáncer de mama y receptores hormonales positivos (HR+) en estadios tempranos con y sin afectación ganglionar.

Figura 2. Diagrama de la tecnología utilizada en la prueba Prosigna™
(http://www.nanosttring.com/media/pdf/Brochure_nCounter_Analysis_System.pdf).



La tecnología en la que se basa el ensayo Prosigna™ es el *nCounter Analysis System*, que consiste en la hibridación directa en solución de los ARNm de la muestra, con sondas específicas de los genes de interés marcadas fluorescentemente con cuatro colores distintos (rojo, amarillo, verde o azul), y analizados en el *nCounter Digital Analyzer*. Los pasos utilizados en esta tecnología se describen en la figura 2. El *nCounter Analysis System* dispone de los perfiles de expresión característicos de los cuatro subtipos moleculares (luminal A, luminal B, HER2 enriquecido y basal), de manera que al introducir los resultados de la expresión de la muestra de un paciente, y mediante un análisis de correlación de Pearson, el sistema asigna el subtipo molecular del paciente de estudio.

Esta tecnología permite su realización en laboratorios locales mediante la adquisición de la instrumentación y los paquetes de reactivos diseñados especialmente para el ensayo. Las características de las dos pruebas genómicas incluidas en este informe se resumen en la Tabla 2.

Tabla 2. Características de las pruebas genómicas EndoPredict® y Prosigna™		
	EndoPredict® (Sividon/Myriad)	Prosigna™ (NanoString Technologies, Inc.)
Genes analizados	12 genes: 3 genes de ciclo de proliferación celular y 5 de señalización hormonal; 4 genes control	50 genes para la identificación del subtipo molecular y 8 genes control
Técnica analítica	RT-PCR sondas taqman	Sondas que hibridan directamente con el ARNm en solución (<i>nCounter Analysis System</i>)
Tejido utilizado	Pieza postquirúrgica en parafina	Pieza postquirúrgica en parafina
Indicación	Mujeres pre y postmenopáusicas con cáncer de mama ER+, HER2- en estadios tempranos con y sin afectación ganglionar	Mujeres postmenopáusicas con cáncer de mama y receptores hormonales positivos, en estadios tempranos con y sin afectación ganglionar
Medida de resultado	Metástasis a 10 años con tratamiento hormonal adyuvante	Metástasis a 10 años con tratamiento hormonal adyuvante
Resultado prueba	Índice EP: basado en los niveles de expresión génica Índice EPclin: algoritmo que combina datos genómicos con los clínicos (tamaño del tumor y estado de afectación ganglionar)	Subtipos moleculares Índice ROR: algoritmo que combina datos genómicos con clínicos y está basado en el subtipo molecular (PAM50), tamaño del tumor, estado de proliferación del tumor y el estado de afectación ganglionar
Clasificación pacientes según los índices	EP: valor de 0–15. Clasifica en bajo y alto riesgo de recaída con un valor de corte igual a 5 EPclin: valor de 1–6,5. Clasifica en bajo y alto riesgo de recaída con hormonoterapia considerando un valor de corte igual a 3,3	Valor de 0–100 clasificado en tres categorías de riesgo: alto, bajo e intermedio Los puntos de corte entre los tres grupos de riesgo son diferentes en función de si hay o no afectación ganglionar
Aplicación clínica	Pronóstica	Pronóstica
FDA/CE	-/con CE	Aprobado/con CE

Para valorar la eficacia en este tipo de pruebas pronóstico es importante establecer y definir la validez analítica y la validez clínica de la prueba así como su utilidad clínica. Según Sanderson *et al.*, 2005:

La validez analítica: indica la exactitud o precisión de la medida de interés, así como de su fiabilidad. En este caso, la medida de interés es la expresión de los ARNm de las células tumorales.

La validez clínica: es la capacidad de la prueba para identificar de forma precisa y fiable a los pacientes o para predecir los resultados clí-

nicos de interés. Normalmente, se presenta en términos de sensibilidad y especificidad clínica.

La utilidad clínica: muestra el balance entre los beneficios y riesgos asociados a la prueba. Indica si la prueba dará lugar a una mejora en los resultados en salud del paciente. Es decir, la prueba no solo debe modificar el manejo clínico, sino que debe traducirse en mejora de la salud del paciente.

Este informe de evaluación ha sido realizado a petición de la Comisión de Prestaciones, Aseguramiento y Financiación, dependiente del Consejo Interterritorial del Sistema Nacional de Salud, a propuesta de la Consejería de Igualdad, Salud y Política Social de la Junta de Andalucía.

Objetivo

El objetivo principal de este trabajo es valorar la eficacia (en términos de validez analítica y clínica) y la utilidad clínica de las pruebas genómicas EndoPredict® y Prosigna™ como pruebas pronóstico del cáncer de mama en estadios tempranos, frente a los índices pronósticos convencionales de carácter clínico-patológico.

2. Material y Métodos

2.1 Tipo de estudio

Se ha realizado una revisión sistemática de la literatura, sobre la eficacia y utilidad clínica de las pruebas genómicas EndoPredict® y Prosigna™ para el pronóstico de pacientes con cáncer de mama ER+/HER2- en estadios tempranos.

2.2 Búsqueda bibliográfica: bases de datos y estrategia

Se ha llevado a cabo una búsqueda bibliográfica en las bases de datos *Cochrane Library*, CRD (*Centre for Reviews and Dissemination*) e INHTA con el fin de identificar estudios de síntesis como revisiones sistemáticas o informes de evaluación de tecnologías (HTA). Además, se realizaron búsquedas tanto de revisiones sistemáticas como de estudios primarios en las bases de referencias MEDLINE (mediante OVID), preMedLine (mediante PubMed), EMBASE y *Web of Science* hasta octubre de 2014. Estas búsquedas se complementaron con la consulta en el *National Institute for Health and Care Excellence* (NICE), en la *Canadian Agency for Drugs and Technologies in Health* (CADTH) y en páginas web de Agencias y Unidades de Evaluación de Tecnologías Sanitarias españolas. Las principales estrategias diseñadas para estas búsquedas se muestran en el Anexo 1. Las búsquedas se han realizado de forma independiente para cada prueba genómica.

Además, la búsqueda bibliográfica se completó con una revisión secundaria de la bibliografía de los artículos obtenidos en las estrategias antes descritas, y búsqueda manual en sitios web de congresos relacionados con la materia.

2.3 Selección de artículos relevantes

Los artículos relevantes que podían dar respuesta a los objetivos de este informe fueron seleccionados siguiendo los siguientes criterios de inclusión y exclusión.

Los criterios de inclusión:

- Población: pacientes con cáncer de mama en estadios tempranos con receptores hormonales positivos.

- Intervención: las pruebas genómicas EndoPredict® o Prosigna™.
- Comparación: con cualquier índice pronóstico clínico-patológico utilizado en la práctica clínica (ej. Índice pronóstico de Nottingham, *St. Gallen Expert Criteria, Adjuvant! OnLine...*).
- Resultados:
 - Validez analítica: reproducibilidad, precisión,
 - Validez clínica: recurrencia de la enfermedad, mortalidad, calidad de vida relacionada con la salud: SF-36, EORTC .
 - Impacto: cambios en la terapia asignada.
- Diseño: Estudios de síntesis basados en la evidencia como revisiones sistemáticas, informes de evaluación de tecnologías sanitarias (HTA) y guías de recomendaciones. En los casos de la existencia de varias RS, informes de evaluación o guías de recomendaciones sobre el tema con las mismas indicaciones e intervenciones, se incluiría la más reciente de mayor calidad (según los criterios establecidos por la guía AMSTAR), evitando de esta manera la redundancia de información. Estudios clínicos con cualquier tipo de diseño que aporte información relevante de la capacidad pronóstico/predictiva de las pruebas.

Los criterios de exclusión:

- Estudios que comparasen las pruebas genómicas (EndoPredict® o Prosigna™) con otros índices genómicos o índices no habituales en la práctica clínica.
- Estudios que no presentasen datos desagregados.
- Estudios a propósito de un caso.
- Comunicaciones a congresos.
- Publicaciones como cartas, editoriales, revisiones no sistemáticas y todas aquellas que no incluyeran datos originales sobre el tema.
- Estudios en idiomas distintos al castellano, inglés, francés e italiano.

2.4 Evaluación de la calidad de los estudios

Se ha realizado una lectura crítica de los artículos seleccionados, con el fin de identificar los problemas metodológicos que pudieran influir en la validez interna y externa de los estudios. En los estudios de validez clínica, la evaluación de la calidad de los estudios de pronóstico se tuvieron en consideración,

a la hora de detectar los posibles sesgos, los criterios establecidos para este tipo de estudios por Hayden, Cote & Bombardier, 2006 y adaptados por el NICE. Los puntos establecidos en estos criterios se centran en la población, el seguimiento, la medida o determinación del factor pronóstico, la medida de los resultados clínicos y los análisis estadísticos realizados (Anexo 2).

2.5 Nivel de evidencia

Los niveles de evidencia se establecieron siguiendo los criterios definidos por la SIGN, (*Scottish Intercollegiate Guidelines Network*) están basados en el tipo de diseño y en el riesgo de sesgos de los estudios. La escala está disponible en: <http://www.sign.ac.uk/pdf/sign50.pdf>.

2.6 Extracción y síntesis de los resultados

Se ha realizado la extracción de los principales resultados descritos en cada estudio y se han resumido en tablas de evidencia siguiendo un modelo desarrollado *a priori*. Se ha realizado una síntesis cualitativa de los datos de los estudios incluidos.

3. Resultados

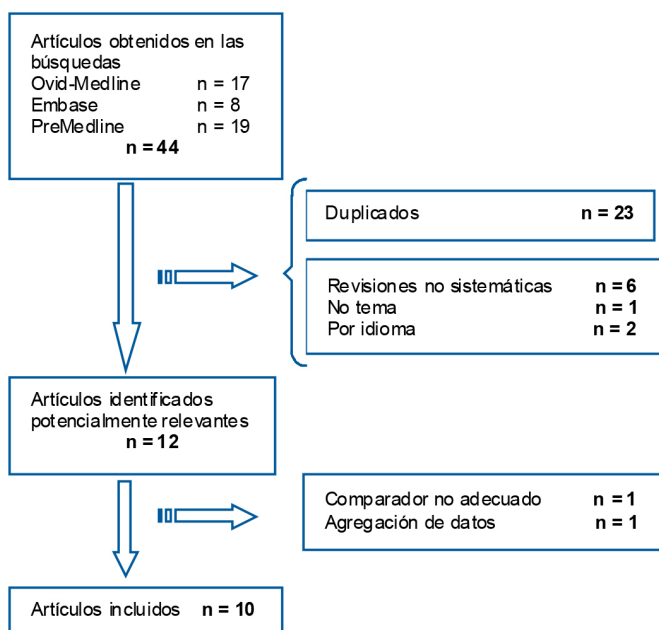
3.1 Resultados EndoPredict®

3.1.1 Resultados de la búsqueda

Las estrategias de búsqueda en las bases de datos referenciales identificaron un total de 44 artículos, de los cuales, 23 fueron descartados por tratarse de duplicados encontrados en las diferentes bases de datos.

Tras la lectura del título y el resumen de las referencias, se excluyeron aquellas que no cumplían los criterios de inclusión establecidos. Se seleccionaron para su lectura a texto completo un total de 12 estudios de los que finalmente, solo 10 han sido incluidos en este informe. El esquema con el flujo de selección de referencias se describe en la Figura 3.

Figura 3. Diagrama de flujo de selección de referencias



Los dos estudios excluidos tras su lectura completa fueron: Varga *et al.*, 2013 y Zhao *et al.*, 2014. El primer estudio se excluyó al tratarse de una comparación entre EndoPredict® y Oncotype DX®, donde estimaron la concordancia y correlación entre ambas pruebas, por lo que no cumplía los criterios de

inclusión al tratarse de un índice genómico no habitual en la práctica clínica. El segundo estudio, Zhao *et al.*, 2014, se excluyó por presentar el análisis estadístico de datos procedentes de un registro donde existían poblaciones tanto ER+ como ER- , con y sin tratamiento, por lo que no cumplía los criterios de inclusión ni por población ni por agregación de datos.

3.1.2 Descripción de los estudios y evaluación de la calidad metodológica

Las 10 referencias identificadas en la búsqueda bibliográfica recogían estudios que aportaban evidencia sobre diferentes aspectos de la prueba: validez analítica, validez clínica e impacto en las decisiones terapéuticas. Los estudios localizados y su finalidad se recogen en la Tabla 3.

Tabla 3. Estudios primarios publicados sobre EndoPredict®			
Referencia	Tipo de estudio	n	Finalidad
Denkert <i>et al.</i> , 2012	Preclínico	10	Validación analítica
Kronenwett <i>et al.</i> , 2012	Preclínico	10	Validación analítica
Müller <i>et al.</i> , 2012	Preclínico	80	Validación analítica
Poremba <i>et al.</i> , 2014	Preclínico	-	Validación analítica
Filipits <i>et al.</i> , 2011	Cohortes retrospectivo procedentes de un ensayo clínico aleatorizado	1.702	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+, HER2- tratadas con terapia hormonal adyuvante
Dubsky <i>et al.</i> , 2013a	Cohortes retrospectivo procedentes de un ensayo clínico aleatorizado	1.702	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+, HER2- (5 años) tratadas con terapia hormonal adyuvante
Dubsky <i>et al.</i> , 2013b	Cohortes retrospectivo procedentes de un ensayo clínico aleatorizado	1.702	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+, HER2- (10 años) tratadas con terapia hormonal adyuvante. Comparación con otros índices clínico-patológicos
Müller <i>et al.</i> , 2013	Retrospectivo	167	Impacto en la decisión terapéutica
Bertucci, Finetti, Viens & Birnbaum, 2014	Retrospectivo	553	Validación clínica (carácter predictivo) en pacientes ER+, HER2- tratadas con quimioterapia neoadyuvante
Martin <i>et al.</i> , 2014	Cohortes retrospectivo procedentes de un ensayo clínico aleatorizado	1.246	Validación clínica (carácter pronóstico) en pacientes ER+, HER2- con afectación ganglionar y tratadas con quimioterapia adyuvante
n: número de pacientes o muestras estudiadas			

Todos los estudios localizados sobre EndoPredict® han sido estudios preclínicos o retrospectivos, no identificándose en la búsqueda ningún estudio de carácter prospectivo. Los artículos de Filipits *et al.*, 2011; Dubsky *et al.*, 2013a y Dubsky *et al.*, 2013b corresponden al mismo estudio pero presentaban análisis complementarios de los resultados, por lo que los tres han sido incluidos en el informe.

Estudios de validez analítica de EndoPredict®

Kronenwett *et al.*, 2012: ha analizado la validez analítica realizando pruebas para determinar la precisión y reproducibilidad de los ensayos dentro de un mismo laboratorio. El estudio valoró en primer lugar, los materiales utilizados y los controles necesarios para lo cual, determinaron los límites de detección analítica, el rango lineal (rango dentro del intervalo de trabajo donde los resultados son proporcionales a la concentración de la sustancia que se está midiendo presente en la muestra), el rango de ARNm necesario en el análisis para evitar cambios en el índice EP y la eficiencia de la PCR, en 12 ensayos de PCR usando diluciones seriadas de las muestras. También valoró la precisión de los resultados utilizando 3 muestras procedentes de tumores diferentes, clasificados según el índice EP como de “bajo riesgo”, de “alto riesgo” o próximos a “alto riesgo”. En el análisis de precisión se comprobaron un amplio número de variables: días ($n = 11$), hora del día ($n = 2$), instrumento de PCR ($n = 4$), posición de la muestra en la placa de 94 celdillas de EndoPredict® ($n = 2$), lote de la placa de EndoPredict® ($n = 4$), lote de la enzima/mezcla de reacción ($n = 2$), y operador ($n = 3$). Además, hicieron un estudio de correlación entre los análisis realizados en dos laboratorios diferentes, el laboratorio donde se ha realizado la validación y un laboratorio independiente de patología molecular. En esta comparación entre laboratorios utilizaron muestras de 10 tumores. El análisis estadístico del estudio consistió en la determinación de los índices EP con sus correspondientes intervalos de confianza (95 %) y la estimación del coeficiente de correlación de Pearson (R^2) para comparar los resultados de las pruebas en dos laboratorios.

Valoración de la calidad metodológica: aunque el estudio ha seguido una adaptación de los criterios estandarizados de verificación y validación de ensayos cualitativos y semicuantitativos multiplex de ácidos nucleicos (Wilson *et al.*, 2008), es importante tener en consideración que estos criterios no incluyen la validación de ensayos de expresión génica ni ensayos cuantitativos como el que se está evaluando en el estudio y en este informe. La preparación de las muestras de ARN se realizó mediante el método automatizado basado en la utilización de perlas magnéticas recubiertas de sílice (VERSANT Tissue Preparation System), lo que es importante a la hora de la validación analítica ya que la extracción de ARN es un paso fundamental y necesario, por lo que, en caso de utilizar otro sistema debe de evaluarse de

nuevo la validez. El estudio está bien diseñado en cuanto al número mínimo de muestras utilizadas, en el caso de la puesta a punto de los materiales se realizaron 12 reacciones de PCR y en el estudio de la precisión de la prueba se usaron al menos 3 tumores diferentes; los análisis se realizaron por triplicado para las determinaciones cuantitativas de los mismos (cálculo de los índices EP). El análisis ha sido riguroso con el fin de evitar datos “outlier” que pudieran dificultar la interpretación de resultados. Los autores presentan los datos desglosados por cada uno de los genes analizados mostrando los resultados adecuadamente y con sus correspondientes intervalos de confianza al 95% lo que permite valorar la precisión de los mismos.

Denkert *et al.*, 2012: ha valorado la validez analítica determinando la reproducibilidad y fiabilidad de la prueba en siete laboratorios diferentes. Para ello, secciones de 10 µm de 10 tumores de mama parafinados se enviaron a 7 laboratorios de patología molecular: 5 laboratorios alemanes, 1 laboratorio austriaco y 1 laboratorio suizo (*Institute of Pathology, Charité Hospital, Berlin; Institute of Pathology, Gelsenkirchen, Germany; Institute of Pathology, Medizinische Hochschule Hannover, Germany; Institute of Pathology, University of Heidelberg, Germany; Institute of Pathology and Pathological Anatomy, Technical University of Munich, Germany; Institute of Pathology, Medizinische Universität Wien, Vienna, Austria; y el Institute of Surgical Pathology, University Hospital Zurich, Switzerland*). Los 10 tumores analizados presentaban un rango en el contenido de células tumorales entre 30–80%, siendo el contenido medio de células tumorales del 63 % y presentaban valores del índice EP entre 2,7 y 12,1. Los tumores eran de grado 1-3, pT1- 3 y pN0-pN3a. Para cada tumor se establecieron unos valores de referencia de EP en los laboratorios de Sividon con los que se compararon los valores obtenidos por cada laboratorio. Todos los laboratorios utilizaron para la extracción de ARN y RT-PCR los mismos instrumentos y reactivos, y siguiendo el mismo protocolo. Se extrajo ARN de una sección de cada uno de los 10 tumores en cada laboratorio y realizaron las PCR por triplicado para poder estimar la media y la imprecisión de la PCR. El estudio tuvo como objetivo primario determinar el número de laboratorios capaces de implementar la prueba de manera satisfactoria. Para ello, se consideró una implementación satisfactoria cuando la diferencia entre los valores locales del EP y el valor de referencia fuera inferior a 1,0 unidad del EP en al menos 9 de las 10 muestras. El análisis estadístico incluía la estimación del coeficiente de correlación de Pearson entre los EP de cada centro con el EP de referencia, la desviación estándar entre el total de las 80 medidas de los 10 tumores realizadas (70 en los centros y 10 de referencia) y el grado de acuerdo en la clasificación del grupo de riesgo según el índice EP mediante un índice de Kappa.

Valoración de la calidad metodológica: el estudio han sido llevado a cabo de manera rigurosa con un buen diseño para validar la reproducibilidad de la prueba en laboratorios diferentes. Se han comparado las pruebas en 7 laboratorios diferentes, siendo laboratorios de 3 países distintos (la mayoría de los laboratorios eran alemanes) y de ámbitos diferentes (hospitalario, universitario...) con el propósito de dar mayor validez externa a la prueba. El análisis estadístico de los datos ha sido adecuado, aunque no está descrito en los materiales y métodos del estudio sino en resultados donde especifican los estadísticos utilizados.

Müller *et al.*, 2012: ha realizado la prueba en dos tipos de muestras diferentes, biopsias y secciones de piezas quirúrgicas. El estudio ha valorado si la forma de obtención del tejido alteraba los resultados y de esta manera, validar la prueba en ambos tipos de muestra. En el proceso de obtención del material biopsiado pueden producirse daños en el tejido, lo que en principio podría afectar al resultado, y de ahí el interés de los autores en esta comparación. Los autores utilizaron parejas de muestras, biopsia-pieza quirúrgica, procedentes de 40 pacientes con tumores de mama ER+, HER2- y determinaron la extracción de ARN y el índice de riesgo EP. Las muestras fueron seleccionadas por presentar cambios inflamatorios inducidos por el proceso de biopsia en secciones de las piezas quirúrgicas teñidas con hematoxilina-eosina (H-E). La mayoría de las biopsias se realizaron bajo control por ultrasonidos y agujas 14G, el tiempo medio entre biopsia y extracción quirúrgica fue de 14 días. Se analizaron 2 secciones consecutivas de 5 µm de cada biopsia y de su correspondiente pieza quirúrgica, una sección fue teñida con H-E y la otra analizada a nivel molecular. Los análisis moleculares se realizaron por triplicado y se calculó el índice EP de cada sección. Se estimó el coeficiente de correlación de Pearson entre biopsia y pieza quirúrgica.

Valoración de la calidad metodológica: el estudio ha valorado un número considerable de pacientes/tumores (40) y se realizó la cuantificación mediante RT-PCR de los 8 genes del EndoPredict® por triplicado en cada una de las muestras. Los resultados presentan un tratamiento estadístico adecuado; los autores no especifican las cantidades de ARN obtenidas de las muestras, aunque comentan que la cantidad de ARN aislada de biopsia es inferior a la obtenida de secciones de la pieza quirúrgica. Este dato es importante ya que según los resultados del estudio de validez analítica de Kronenwett *et al.*, 2012 existe un rango de ARN necesario en el análisis para evitar cambios en el índice EP. Los autores comentan que la cantidad fue suficiente pero deberían de mostrarse en el estudio; así como los valores de cada una de las parejas de muestras, con su intervalo de confianza para poder comprobar la precisión de las medidas. Sólo presentan el análisis final de correlación.

Poremba *et al.*, 2014: han estudiado si las condiciones de análisis afectaban el resultado de la prueba EndoPredict®, es decir, a su reproducibilidad y su fiabilidad. En concreto analizaron variaciones en 4 parámetros preanalíticos relevantes ya que dichos parámetros pueden afectar a la calidad del ARN y por lo tanto afectar a los resultados de la prueba, al tratarse de una prueba basada en ARN. Los parámetros o variables analizadas han sido: tiempo y temperatura de almacenamiento de la muestra hasta su fijación, tiempo del proceso de fijación, contenido de células tumorales en la muestra y tiempo de almacenamiento de la muestra hasta su utilización. Se utilizó una sola pieza quirúrgica que se dividió en 9 muestras de igual tamaño y fueron las utilizadas en el estudio. En total realizaron 138 ensayos de EndoPredict® entre los realizados para cada una de las variables. Los autores han estimado las variaciones medias de las medidas y los coeficientes de correlación de Pearson para comparar parejas de muestras con distintos contenidos de células tumorales.

Valoración de la calidad metodológica: aunque realiza en total 138 ensayos, eso no significa que se haya realizado un número tan elevado de ensayos en el análisis de cada variable. En el análisis del impacto del tiempo hasta la fijación de la muestra en realidad estudian dos variables: el tiempo hasta la fijación y la temperatura a la que se encuentra la muestra durante ese tiempo. El tiempo medido fue 10 min, 1 h, 12 h y 24 h y la temperatura fue 4°C, 20°C y 37°C, no teniendo muestras de todas las situaciones posibles. De cada una de las muestras, se realizaron extracciones de ARN de 2 secciones de 10 µm, lo mínimo serían 3 secciones. Por lo tanto, realizan los cálculos estadísticos basados en 2 extracciones de ARN de cada muestra y las PCR de cada muestra se hicieron por triplicado. De hecho, la muestra que estuvo 24h a 20°C hasta su fijación no presenta cambios en cuanto al proceso de expresión del gen utilizado como marcador subrogado (gen *RPL37A*) de la cantidad de ARNm, pero sí presenta cambios relevantes en el índice EP, que es el que clasifica el riesgo de los pacientes. Al haber sólo una muestra en el estudio, no se puede saber si es un valor fiable o no. Esta misma situación de falta de repetitividad en diferentes muestras ocurre en la variable del tiempo de fijación, con una única muestra dividida en 6 secciones y cada sección analizada a un tiempo. En general, los resultados no tienen la robustez estadística necesaria para que sean generalizables y por lo tanto, concluyentes.

Los cuatro estudios, por consiguiente, analizaron la reproducibilidad de la prueba y los posibles factores que pueden influir en dicha reproducibilidad. Estos estudios de reproducibilidad son muy importantes al tratarse de una prueba genómica diseñada para realizarse de forma descentralizada en aquellos laboratorios donde interese incorporarla.

Estudios de validez clínica de EndoPredict®

Los estudios de validez clínica centrados en demostrar la capacidad de la prueba para identificar de forma precisa y fiable el riesgo de recurrencia a distancia de la enfermedad o para predecir los resultados clínicos de interés, se han llevado a cabo en estudios retrospectivos de cohortes que habían sido evaluadas de forma prospectiva al estar incluidas en ensayos clínicos aleatorizados. Los estudios han analizado la prueba EndoPredict® en muestras tumorales disponibles procedentes de pacientes incluidas en ensayos clínicos aleatorizados realizados previamente y diseñados para comparar diferentes tratamientos de terapia hormonal. Los ensayos clínicos que han aportado los pacientes han sido básicamente tres y tenían los siguientes objetivos:

- Ensayo ABCSG-6: estudio realizado en mujeres postmenopáusicas con cáncer de mama ER+ en estadios tempranos con el fin de valorar si el tratamiento combinado durante 2 años de tamoxifeo con aminoglutetimida 500 mg, seguido de tres años con solo tamoxifeno presentaba mejores resultados que un tratamiento de 5 años con solo tamoxifeno.
- Ensayo ABCSG-8: realizado en mujeres postmenopáusicas con cáncer de mama ER+ en estadios tempranos con el objetivo de comparar si dos años de tratamiento con tamoxifeno seguidos de 3 años con anastrozol mejoraba los resultados de supervivencia frente a continuar esos 3 años con tamoxifeno.
- Ensayo GEICAM/9906: este estudio comparaba dos tratamientos de quimioterapia adyuvante en mujeres (pre y postmenopáusicas) con cáncer de mama ER+/HER2- y con al menos 1 ganglio afectado.

La descripción de los estudios de validez clínica y la valoración del riesgo de sesgo de los mismos se ha recogido en la Tabla 4.

Básicamente, los estudios de validación clínica han analizado la recurrencia a distancia de la enfermedad. Para ello, han clasificado a las pacientes de estudio en dos grupos según el riesgo de recaída, grupo de bajo riesgo y de alto riesgo, y han estimado mediante análisis estadísticos de Kaplan-Meier la supervivencia o recurrencia de la enfermedad a 5 y 10 años.

En cuanto a la calidad de la evidencia aportada por el estudio principal sobre la validez clínica de la prueba (Filipits *et al.*, 2011; Dubsy *et al.*, 2013a y Dubsy *et al.*, 2013b), hay que señalar que aunque el diseño de estudio utilizado para la validación clínica de la prueba es de carácter retrospectivo, se puede considerar un diseño adecuado y pertinente al objetivo que se pretende. El cáncer de mama presenta un alto porcentaje de supervivencia a 5 y 10 años. La realización de ensayos prospectivos requeriría un tiempo de

seguimiento demasiado largo para la obtención de resultados concluyentes. En este caso, el hecho de disponer de dos cohortes de pacientes procedentes de ensayos clínicos aleatorizados y seguidas prospectivamente durante más de 5 años, permite valorar con cierto grado de seguridad la validez de la prueba. Aunque en los estudios la prueba EndoPredict® se ha realizado retrospectivamente, existen una serie de aspectos desde el punto de vista metodológico que dan calidad a los estudios. Estos aspectos son importantes de remarcar ya que implican un esfuerzo en controlar los posibles sesgos existentes.

- La población de estudio procede de ensayos clínicos aleatorizados, lo que implica una selección previa de la población con unos criterios de inclusión definidos y estrictos, que reduciría el posible sesgo de selección. Además, aunque los ensayos aleatorizados no estaban diseñados para este estudio en concreto, las variables resultado coinciden con las de interés, lo que garantiza en cierto grado que la recogida de datos de las variables sea correcta y no implique un sesgo importante de realización.
- Las muestras a las que se les ha realizado el EndoPredict® fueron tomadas en el momento de la cirugía antes de la terapia adyuvante y se almacenaron a temperatura ambiente. Todos los análisis moleculares fueron realizados de forma centralizada en un único laboratorio de patología molecular reduciendo los sesgos de realización.
- Los objetivos, la metodología de los ensayos, el procedimiento de cálculo de los índices y los valores de corte fueron preestablecidos.
- Los análisis moleculares se realizaron de forma cegada, no conociendo los analistas ni los datos ni los resultados clínicos.
- El estudio ha realizado un análisis estadístico correcto y exhaustivo con el fin de detectar variables confusoras y controlar los sesgos.

En resumen, la evidencia del principal estudio de valoración clínica del EndoPredict® como factor pronóstico, siguiendo los criterios propuestos por SIGN, mostraba un nivel de evidencia 2+ (estudio de cohortes bien realizado con bajo riesgo de sesgos).

El otro estudio de validación clínica de la prueba realizado por Martin *et al.*, 2014, ha seguido el mismo diseño metodológico que los anteriores pero centrado en una población de mujeres, pre y postmenopáusicas, con afectación ganglionar. La población procede de un ensayo clínico aleatorizado con finalidad terapéutica y como los propios autores describen, el estudio sigue los criterios de calidad propuestos por Simon *et al.*, 2009 para la valoración

de biomarcadores pronóstico en muestras almacenadas. Presenta las mismas fortalezas y debilidades metodológicas descritas para el estudio anterior, por lo que la evidencia que aporta se considera de un nivel de evidencia 2+ (estudio de cohortes bien realizado con bajo riesgo de sesgos).

Los estudios de Müller *et al.*, 2013 y Bertucci *et al.*, 2014 se corresponden a estudios retrospectivos pre-clínicos o de series de casos, por lo que el diseño de los mismos llevan implícitos un alto riesgo de sesgo, aportando un nivel 3 de evidencia (series de casos).

Por último, en cuanto a los posibles conflictos de interés, señalar que en la mayoría de los artículos incluidos algunos de los autores pertenecían a *Sividon Diagnostic* o habían recibido algún tipo de financiación de dicha compañía, quedando explícitos los conflictos de interés en todas las publicaciones.

Tabla 4. Características de los estudios de validez clínica de EndoPredict® incluidos en el informe		
Estudio	Características del estudio	Diseño del estudio
Filipits <i>et al.</i> , 2011 Dubsky <i>et al.</i> , 2013a Dubsky <i>et al.</i> , 2013b	Muestra para el desarrollo matemático de los índices: 964 tumores de mama ER+, HER2 -, procedentes de pacientes con cáncer de mama en estadios tempranos con tratamiento hormonal adyuvante únicamente con tamoxifeno	Diseño: estudio pre-clínico de análisis de la expresión génica y estudio de cohortes retrospectivo procedentes de un ensayo clínico aleatorizado para la validación de los índices de riesgo EP y EPclin
Financiación: Robert Bosch Foundation y Federal Ministry of Education and Research, Germany	Población para la validación clínica de los índices: 1702 mujeres postmenopáusicas con cáncer de mama ER+, HER2 - en estadios tempranos procedentes de dos ensayos clínicos aleatorizados y reclutadas en 24 centros diferentes	Objetivo: desarrollo matemático de los índices EP y EPclin y estudio de validación del carácter pronóstico de los mismos en mujeres postmenopáusicas con cáncer de mama ER+, HER2 - en estadios tempranos sin afectación ganglionar
Conflictos de interés: al menos 8 autores recibían honorarios de varias compañías farmacéuticas, incluida <i>Siwon Diagnostics</i> ; o tenían intereses personales en ellas	ABCSG-6: n= 378 del brazo de tamoxifeno. ABCSG-8: n= 1324 En las dos muestras (964 y 1702) había un 65 % y 68 % de pacientes sin afectación ganglionar	Realización: se analizó la expresión génica en una muestra de 964 tumores y se desarrolló el algoritmo matemático del índice EP basado en una combinación lineal de la expresión relativa de cada uno de los genes de interés y definido en un rango de 0 a 15 para evitar valores negativos. El EPclin se desarrolló combinando el valor del EP y dos parámetros clínicos mediante la siguiente fórmula: EPclin= 0,35 t + 0,64 n +0,28 EP - t: es el tamaño de tumor (1: ≤1 cm, 2: >1 y ≤2 cm, 3: >2 y ≤ 5 cm, 4:> 5 cm) - n: el número de ganglios afectados (1: sin afectación ganglionar, 2: 1-3 ganglios positivos, 3: 4-10 ganglios positivos y 4: > 10 ganglios positivos)
Calidad metodológica: moderado/bajo riesgo de sesgo Nivel de evidencia: 2+	Clasificación de los pacientes mediante dos índices pronóstico: <i>Adjuvant! Online St. Gallen, MCCN, German S3 y EP/EPclin</i>	Una vez desarrollados los índices se validaron en una cohorte de 1702 pacientes. Los pacientes fueron clasificados según los dos índices en pacientes con alto o bajo riesgo y se analizó la recurrencia mediante análisis Kaplan-Meier
	Medidas de resultado: recurrencia a distancia de la enfermedad a los 5 y 10 años	Seguimiento: ensayo ABCSG-6: 97,4 meses. Ensayo ABCSG-8: 72,3 meses
		Limitaciones: estudio de cohortes retrospectivo, el mayor riesgo proviene del sesgo de memoria que afectaría a las posibles variables confusoras que pudieran afectar en el estudio actual y no fueron recogidas en su momento. Los autores comentan que una limitación de su estudio era el escaso número de tumores de grado 3 en relación a lo que se esperaba en mujeres post-menopáusicas

Tabla 4. Características de los estudios de validez clínica de EndoPredict® incluidos en el informe (continuación)		
Estudio	Características del estudio	Diseño del estudio
Martin <i>et al.</i>, 2014 Financiación: - Conflictos de interés: al menos 6 autores recibían honorarios de <i>Sivodon Diagnostics</i>, o tenían intereses personales en la compañía Calidad metodológica: moderado/ bajo riesgo de sesgo Nivel de evidencia: 2+	Población: n= 800 mujeres (18-75 años) con cáncer de mama ER+, HER2 - en estadios tempranos con al menos un ganglio afectado y tratadas con quimioterapia adyuvante seguida de terapia hormonal. Las pacientes procedían del ensayo clínico aleatorizado GEICAM/9906 Clasificación de los pacientes mediante dos índices pronóstico: EP/EPclin Medidas de resultado: - Supervivencia libre de recurrencia a distancia de la enfermedad (tiempo entre la aleatorización hasta la muerte o recurrencia a distancia) - Supervivencia global	Diseño: estudio de cohortes retrospectivo procedentes de un ensayo clínico aleatorizado para la validación de los índices de riesgo EP y EPclin Objetivo: validación clínica del carácter pronóstico de los índices EP y EPclin en mujeres con cáncer de mama ER+, HER2 - en estadios tempranos con afectación ganglionar Realización: los objetivos, los valores de corte y el análisis estadístico se especificaron de manera previa al estudio, siguiendo las recomendaciones dadas para estudios de biomarcadores predictivos con muestras almacenadas (Simon, Paik & Hayes, 2009) Los puntos de corte para la clasificación del grupo de riesgo fueron: - Bajo riesgo: EP < 5 (EPclin < 3,3) - Alto riesgo: EP ≥ 5 (EPclin ≥ 3,3) Análisis estadístico de Kaplan-Meier para ver la supervivencia de cada grupo de riesgo y la prueba <i>Log-Rank</i> para comparar las variables entre grupos: HR (<i>hazard ratio</i>) y sus intervalos de confianza al 95 % para cada variable mediante un análisis multivariante de COX Seguimiento: 8,7 años Limitaciones: estudio retrospectivo por lo que el mayor riesgo proviene del sesgo de memoria Con el índice EP en los resultados se presentaban la HR (IC 95 %) y la reducción del riesgo absoluto (RRA) de los dos grupos de riesgo, pero en el análisis realizado con el índice EPclin sólo se aporta el RRA, no mostrando la HR ni los intervalos de confianza de los estadísticos. Esta presentación de los resultados no permite valorar la precisión del índice EPclin lo que limitaría la interpretación del índice

Tabla 4. Características de los estudios de validez clínica de EndoPredict® incluidos en el informe (continuación)		
Estudio	Características del estudio	Diseño del estudio
Bertucci <i>et al.</i>, 2014 Financiación: Institut Paoli-Calmatte, Inserm, Institut National du Cancer (AOPL2010 VOIRES), Ligue Nationale contre le Cancer (label DB) y SIRIC INCa-DGOS-Inserm 6038 Conflictos de interés: no Calidad metodológica: alto riesgo de sesgo Nivel de evidencia: 3	Población: n= 553 mujeres (mediana de edad 49 años) con cáncer de mama ER+, HER2 - en estadios tempranos tratadas con quimioterapia neoadyuvante con antraciclinas Clasificación de los pacientes mediante dos índices pronóstico: EP/EPclin y su respuesta a tratamiento con quimioterapia neoadyuvante Medidas de resultado: - Respuesta a tratamiento quimioterápico neoadyuvante (pCR), considerando: - pCR: grado 1–2 en pieza quirúrgica - no-pCR: grado 3–4 en pieza quirúrgica Supervivencia libre de recurrencia a distancia de la enfermedad (tiempo entre la aleatorización hasta la muerte o recurrencia a distancia)	Diseño: estudio de retrospectivo de una serie de casos Objetivo: valorar el carácter predictivo de los índices EP y EPclin en la respuesta al tratamiento con quimioterapia neoadyuvante Realización: análisis univariante y multivariante para la predicción a respuesta a quimioterapia neoadyuvante Análisis estadístico de Kaplan-Meier para ver la supervivencia de cada grupo de riesgo y la prueba <i>Log-Rank</i> para comparar las variables entre grupos Seguimiento: 40 meses Limitaciones: estudio retrospectivo con posibles variables confusoras que pudieran afectar a la interpretación de los resultados. Entre estas variables se encontrarían: las posibles modificaciones en la preparación del material parafinado, almacenamiento de las diferentes muestras. Los distintos tratamientos neoadyuvantes recibidos por las pacientes...

Tabla 4. Características de los estudios de validez clínica de EndoPredict® incluidos en el informe (continuación)		
Estudio	Características del estudio	Diseño del estudio
Müller <i>et al</i>, 2013 Financiación: "University research advancement" Conflictos de interés: 2 autores declaran conflicto de interés al ser accionistas de <i>Sividon Diagnostics</i>. Calidad metodológica: alto riesgo de sesgo Nivel de evidencia: 3	Población: n= 168 muestras tumorales procedentes de mujeres con cáncer de mama ER+, HER2 - entre agosto 2011 y julio 2012 Medidas de resultado: - Tiempo medio de realización de la prueba - Cambio en la decisión terapéutica Correlación entre el índice EP y el grado tumoral; y entre EP y la proliferación celular	Diseño: realización de la prueba EndoPredict® de muestras retrospectivas en un laboratorio y estudio de encuestas, antes-después de la prueba, a los profesionales Objetivo: comprobar la realización de la prueba dentro de la práctica clínica y valorar el impacto sobre decisión terapéutica Realización: se realizó durante un año la prueba en 168 muestras tumorales para su valoración como prueba rutinaria en el laboratorio. Se realizó una comparación de la decisión terapéutica antes y después del resultado del índice EPclin. La valoración retrospectiva de la decisión terapéutica se realizó mediante una encuesta formada por dos preguntas; antes y después de conocer EPclin y con dos opciones de respuesta cada una: terapia hormonal o terapia hormonal + quimioterapia. Analizaron mediante métodos estadísticos la correlación de los índices con el grado tumoral y la proliferación celular Limitaciones: estudio retrospectivo con posibles variables confusoras que pudieran afectar a la interpretación de los resultados. Además, en la valoración del impacto en la decisión terapéutica, las encuestas presentan limitaciones como es la falta de las señales o información visual en los registros y la complejidad de la situación "in vivo", además, la evaluación de los datos siempre depende de la calidad los registros (cantidad de información recogida y la calidad de la misma)

3.1.3. Resultados clínicos de EndoPredict®

Validez analítica

Es difícil de evaluar la validez analítica del EndoPredict® al no existir actualmente una prueba patrón de referencia para estimar la expresión génica. Los estudios disponibles se han centrado en analizar la reproducibilidad de la prueba, tal como se ha realizado con las plataformas genómicas de primera generación Oncotype DX® y MammaPrint®.

El estudio principal de validación analítica, Kronenwett *et al.*, 2012, ha establecido las condiciones de experimentación de la RT-PCR y su reproducibilidad. Los resultados del estudio se resumen en la Tabla 5.

Tabla 5. Resultados del estudio de validación analítica de EndoPredict® de Kronenwett <i>et al.</i> , 2012	
Variable	Resultado
Límite de blanco	40 Cq
Límite de detección	35,1–37,2 Cq
Rango lineal	35,1–37,2 Cq
Eficiencia de PCR	76 –101%, media 88 %
Rango de concentración ARN	0,16 –18,5 ng/ul
Precisión	SD = 0,15 (1% del rango total del EP) la mayor parte del SD; SD = 0,14 fue debida a la ruido en el proceso de replicación en la PCR y no a condiciones de operador
Reproducibilidad entre 2 laboratorios	Pearson (R2) = 0,995
Cq: unidades de cuantificación, correspondientes a ciclos de PCR SD: desviación típica	

El estudio de Denkert *et al.*, 2012 mostró que la desviación típica media de las 80 medidas realizadas era de 0,25 unidades de EP y que en el 86 % de las muestras la diferencia entre el valor local y el de referencia era de 0,5 unidades de EP. Se observó una buena correlación entre los valores locales de cada uno de los 7 laboratorios con el de referencia, en un rango coeficiente de correlación de Pearson entre 0,987 y 0,999. El coeficiente de Pearson global para todos los valores comparados con los de referencia fue de 0,994. Todas las muestras fueron clasificadas correctamente en su grupo de riesgo con una concordancia del 100% y un índice de Kappa de 1.

La variables preclínicas valoradas en el estudio de Poremba *et al.*, 2014 no parecían afectar de manera general a los resultados del índice EP de EndoPredict®, al igual que tampoco parece afectar la estimación del índice EP si la muestra procede de biopsia o de pieza quirúrgica (Müller *et al.*, 2012). Si bien, se observó que el índice EP de clasificación disminuía cuando la muestra se mantenía a 20°C durante 24h antes de su fijación y cuando en el proceso de fijación se incubaba en formalin 10 días, observándose además, que la cantidad de ARN que se aislaba de biopsia era mucho menor que la obtenida de secciones de 5 µm de espesor de la pieza quirúrgica. Estos resultados han mostrado por lo tanto, que es importante establecer un protocolo de experimentación que evite posibles variaciones en la valoración del EP.

Es importante señalar, que en los estudios de validez analítica los laboratorios implicados son laboratorios de patología molecular, es decir laboratorios que cuentan con un cierto grado de especialización y experiencia en este tipo de técnicas de biología molecular. La representatividad de la muestra de laboratorios para asumir la validez externa del procedimiento podría no ser la más adecuada, por lo que sería conveniente valorar si estos laboratorios son iguales a los disponibles en nuestro sistema sanitario.

El estudio de Müller *et al.*, (2013) realizó de forma rutinaria durante un año la prueba EndoPredict® en el laboratorio de patología molecular del Hospital Universitario Charité de Berlín, y mostraron que la mediana del tiempo medio de realización de la prueba fue de 3 días (rango: 0 a 11 días). El 59,3 % de las muestras se realizaron en un tiempo igual o inferior a 3 días. Las razones para el retraso en la realización de las pruebas fueron: la pérdida de datos clínicos y la repetición en la extracción de ARN (8 de los 168 casos); en un caso fue necesario un bloque de tejido parafinado nuevo. En otro caso no fue posible realizar la prueba al no conseguir extraer ARN de la muestra.

Validez clínica

Los estudios de validación clínica han analizado la recurrencia a distancia de la enfermedad a 5 y 10 años. Para ello, han clasificado a las pacientes de estudio en dos grupos según el riesgo de recaída, grupo de bajo riesgo y de alto riesgo, y han estimado mediante análisis estadísticos de Kaplan-Meier la supervivencia o recurrencia a 5 y 10 años de cada grupo. En la población de estudio se registraron 160 eventos de recurrencia a distancia en los 10 años de seguimiento.

Filipits *et al.*, (2011) en el desarrollo del índice EPclin encontraron que en los análisis multivariantes, el EP, junto con la afectación glandular y el tamaño de tumor eran los factores que contribuían de manera significativa en la predicción de la recurrencia a distancia en las dos cohortes estudiadas. De ahí, que la fórmula para la estimación del EPclin fuera la siguiente:

$$EP_{clin} = 0,35 t + 0,64 n + 0,28 EP$$

- t: es el tamaño de tumor (1: ≤ 1 cm, 2: > 1 y ≤ 2 cm, 3: > 2 y ≤ 5 cm, 4: > 5 cm).
- n: el número de ganglios afectados (1: sin afectación ganglionar, 2: 1-3 ganglios positivos, 3: 4 – 10 ganglios positivos y 4: > 10 ganglios).

Los resultados sobre validez clínica se resumen en las Tablas 6 y 7. Se muestran las diferencias en cuanto a la recurrencia de la enfermedad a distancia (*Hazard Ratio*) entre los dos grupos de riesgo establecidos por el índice EP (sólo información genómica) y por el índice EP_{clin} (información genómica+clínica) a 5 y ≥ 10 años.

Tabla 6. Resultados de validez clínica de los índices EP y EP_{clin} en mujeres postmenopáusicas ER+, HER-2-

Estudio	Población (n)	EP		EP _{clin}	
		5 años HR (IC95 %)	≥ 10 años HR (IC95 %)	5 años HR (IC95 %)	≥ 10 años HR (IC95 %)
Filipits, 2011	ABCSG-6 (378)	-	3,64 (1,85–7,16)*	-	7,97 (3,56–17,8)*
	ABCSG-8 (1324)	-	2,66 (1,67–4,23)*	-	4,27 (2,74–6,67)*
Dubsky, 2013a	ABCSG-6/8 (1702)	2,80 (1,81–4,34)*	3,28 (1,48–7,24)*	4,82 (3,12–7,44)*	6,25 (2,72–14,36)*

* P (LogRank) $< 0,001$

La valoración de la contribución de las distintas variables clínico-patológicas y moleculares a la predicción de la recurrencia a distancia de la enfermedad, se realizó mediante el análisis de combinación de variables “*c-index*”, y se observó que la incorporación del índice EP en el análisis incrementaba de manera significativa el poder pronóstico y además, el índice EP_{clin} era la variable de mayor índice *c*, siendo $c = 0,788$ en la cohorte ABCSG-6 y $c = 0,732$ en ABCSG-8 (Filipits *et al.*, 2011). Estos análisis mostraban a EndoPredict® como un factor pronóstico independiente que añadía información pronóstica a los índices clínico-patológicos utilizados en la práctica clínica (*Adjuvant! Online* en combinación con Ki67 y ER). Este valor pronóstico para la recurrencia a distancia se observaba tanto a 5 como a 10 años (Dubsky *et al.*, 2013a).

Por su parte el estudio de Martin *et al.*, 2014, realizado en una población de mujeres pre y postmenopáusicas, tratadas con quimioterapia adyuvante, mostró que la estratificación de las pacientes en dos grupos de riesgo, alto y

bajo, según los índices EP y EPclin, presentaba a los 5 años de seguimiento diferencias significativas en la recurrencia de la enfermedad a distancia entre ambos grupos. El índice EP clasificó al 20 % de las pacientes (n = 141) como de bajo riesgo de recurrencia y el EPclin clasificó de bajo riesgo al 13 % (n = 74). Los resultados se resumen en la Tabla 7.

Tabla 7. Resultados de validez clínica de los índices EP y EPclin en mujeres con cáncer de mama ER+, HER-2- tratadas con quimioterapia adyuvante a 10 años (Martin <i>et al.</i> , 2014)		
Variable resultado	EP	EPclin
SLRD	Grupo bajo riesgo: 93 % Grupo alto riesgo: 70 %	Grupo bajo riesgo: 100 % Grupo alto riesgo: 72 %
HR (95 % IC) de SLRD	4,8 % (2,5–9,6%)	-
RAR (95 % IC) de SLRD	23 %, <i>p</i> < 0,0001	28 %, <i>p</i> < 0,0001
OS	Grupo bajo riesgo: 92 % Grupo alto riesgo: 67 %	Grupo bajo riesgo: 99 % Grupo alto riesgo: 69 %
HR (95 % IC) de OR	3,9 % (2,0 –7,5 %)	-
SLRD: supervivencia libre de recurrencia a distancia; OS: supervivencia global; RAR: reducción absoluta de riesgo; HR: <i>hazard ratio</i> ;		

Los pacientes clasificados en el grupo de bajo riesgo según los índices EP y EPclin, mostraron una supervivencia libre de recurrencia y una supervivencia global mayor que los de alto riesgo, con una reducción del riesgo absoluto del 23 y 28 %, respectivamente. Como en el estudio de Filipits *et al.*, 2011, el análisis de diferentes combinaciones de factores de riesgo clínico-patológicos y los índices EP y EPclin mostraron que el EPclin presentaba el índice *c* más alto, *c* = 0,7 y que la incorporación del EP a los parámetros clínicos aumentaba de manera significativa el índice *c*, aunque los valores en esta población eran inferiores a los encontrados en mujeres postmenopáusicas sin quimioterapia de los ensayos clínicos ABCSG-6 y ABCSG-8.

El estudio además, analizaba la clasificación según los índices EP y EPclin en dos subgrupos, por una parte en mujeres pre y postmenopáusicas y por otra parte en los dos brazos de tratamiento del ensayo clínico aleatorizado GEICAM/9906. Tanto en mujeres premenopáusicas como postmenopáusicas, los índices EP y EPclin mostraban diferencias significativas entre los grupos de alto y bajo riesgo en supervivencia libre de recurrencia a distancia; aunque en mujeres premenopáusicas los índices EP y EPclin mostraron mejores resultados (Tabla 8). El estudio no observó diferencias significativas entre los dos grupos de tratamiento.

Tabla 8. Análisis de subgrupos del estudio de Martin *et al.*, 2014 en mujeres con cáncer de mama ER+, HER2- tratadas con quimioterapia adyuvante a 10 años

Variable	EP		EPclin	
	Premenopáusicas	Postmenopáusicas	Premenopáusicas	Postmenopáusicas
HR (IC95 %) de SLRD	6,68 (2,44–18,30) $p < 0,0001$	3,34 (1,32–8,45) $p < 0,0069$	-	-
RAR	26 %	18 %	30 %	24 %

HR: *hazard ratio*; SLRD: supervivencia libre de recurrencia a distancia; RAR: reducción absoluta de riesgo.

Según el índice EP el riesgo de recurrencia del grupo de alto riesgo era el doble en mujeres premenopáusicas (6,68) que en mujeres postmenopáusicas (3,34).

Bertucci *et al.*, 2014, al valorar el valor predictivo de EndoPredict® en la respuesta a la quimioterapia neoadyuvante con antraciclinas (pCR), encontraron que el 81 % (IC 95 % 75–87 %) de las 553 pacientes respondieron al tratamiento. Al valorar la pCR en los dos grupos de riesgo establecidos por el índice EP se observó una correlación entre el índice EP y la respuesta al tratamiento, así pCR era inferior en el grupo de bajo riesgo (7 %, 19/283) que en el grupo de alto riesgo (17 %, 45/270). Esta diferencia entre los dos grupos era estadísticamente significativa ($p < 0,001$).

Los estudios de Dubsky *et al.*, (2013b) y Varga *et al.*, (2013) comparaban la estratificación del riesgo de recurrencia a distancia de los índices de EndoPredict® con la estratificación establecida por otros índices pronóstico, clínico-patológicos y genómicos, respectivamente.

Dubsky *et al.*, (2013b) al comparar EndoPredict® con los índices clínico-patológicos establecidos en: Guía S3 alemana del 2008, *American National Comprehensive Cancer Center Network* (NCCN) del 2007 y *St Gallen* del 2011, en una muestra de 1702 mujeres postmenopáusicas con cáncer de mama ER+, HER2- sin afectación ganglionar; el índice EPclin estratificaba el riesgo de recurrencia a distancia a 10 años, en alto y bajo riesgo, mejor que los otros índices. Los índices clínico-patológicos, la guía Guía S3, la guía NCCN y la guía *St Gallen*, clasificaron de bajo riesgo al 15 %, 6 % y 19 % de las de las 1.702 pacientes, respectivamente; mientras que el índice EPclin clasificó de bajo riesgo al 63 % de las pacientes estudiados. Todos los índices parecen identificar claramente un grupo de pacientes con un riesgo muy bajo de recurrencia de la enfermedad, ya que todos los grupos presentan una supervivencia de las pacientes con bajo riesgo, muy alta a los 10 años (Tabla 9). Para realizar las comparaciones con los índices *St Gallen* y Guía S3 donde la estratificación es en riesgo bajo, intermedio y alto, el riesgo intermedio se une al alto quedando por consiguiente dos grupos: riesgo bajo y riesgo intermedio/alto.

Tabla 9. Comparación de EPclin con los índices clínico-patológicos (Dubsky *et al.*, 2013b)

Índice pronóstico	HR (IC95 %)	<i>P</i> (LogRank)*	RRA (IC95 %)	% pacientes libres recurrencia a 10 años
Guía S3	2,20 (1,16–4,19)	0,014	7,9 % (3,0 % – 12,9 %)	94,7 % (90,5 % – 98,9 %)
NCCN	2,16 (0,80–5,85)	0,12	6,9 % (0,9 % – 13 %)	94,5 % (88,9 % – 100 %)
St Gallen	2,78 (1,50–5,14)	< 0,001	11,2 % (7,7 % – 14,7 %)	96,9 % (94,9 % – 98,9 %)
EPclin	5,11 (3,48–7,51)	< 0,001	18,7 % (13,5 % – 23,9 %)	95,3 % (93,4 % – 97,3 %)

* se considera estadísticamente significativa $p < 0,05$; RRA: reducción del riesgo absoluto a 10 años

En resumen, los estudios que han valorado la validez clínica de EndoPredict® han mostrado que:

- Los índices EP y EPclin clasificaban los pacientes en grupos de alto y bajo riesgo de recurrencia a distancia, y que existían diferencias significativas entre ellos a 5 y 10 años tanto en mujeres pre y post menopáusicas con o sin afectación ganglionar.
- EndoPredict® parecía clasificar mejor el riesgo de recurrencia que los índices Guía alemana S3 del 2008, NCCN del 2007 y *St Gallen* del 2011.

Utilidad clínica

No se ha identificado ningún estudio con el diseño adecuado para poder valorar la utilidad de la prueba EndoPredict®, ya que para validar o establecer la utilidad clínica se necesitarían estudios que valoraran de manera prospectiva la recurrencia de la enfermedad teniendo en consideración los índices EP y EPclin en la decisión terapéutica.

Impacto en la decisión clínica

Solamente se ha localizado un estudio que valorara el impacto de la prueba en la decisión terapéutica, Müller *et al.*, 2013. El estudio incluía 167 muestras con resultados de pruebas de EndoPredict® aunque sólo se dispuso de información terapéutica en 130 de los casos (77,8 %).

Antes de la prueba, 47 pacientes (36,2 %) se consideraron para terapia hormonal y 83 pacientes (63,8 %) se consideraron para terapia combinada (terapia hormonal más quimioterapia); tras conocer los resultados de EndoPredict® el número de pacientes considerados para terapia hormonal pasó a 67 (51,5 %) y solamente 62 (47,7 %) fueron considerados para terapia

combinada. La prueba modificaba la decisión terapéutica en el 37,7 % de los pacientes y reducía el número de pacientes tratados con quimioterapia.

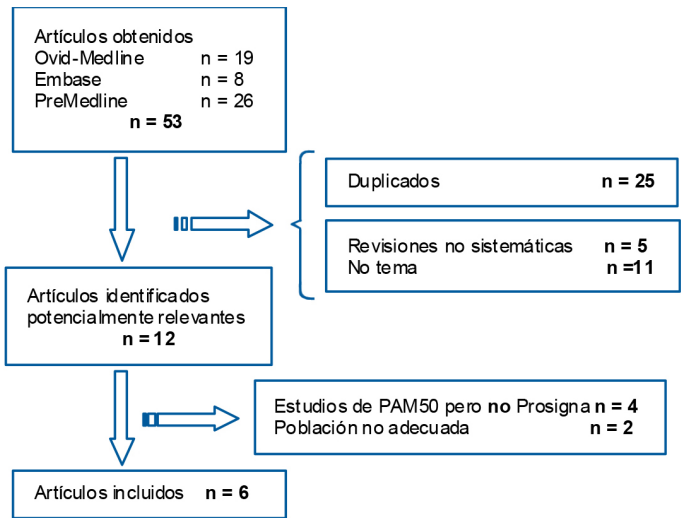
3.2 Resultados Prosigna™

3.2.1 Resultados de la búsqueda

Las estrategias de búsqueda en las bases de datos referenciales identificaron un total de 53 artículos. Del total de referencias encontradas, 25 se descartaron por tratarse de duplicados encontrados en varias bases de datos. La lectura del título y el resumen de las referencias, permitió la selección de 12 estudios que podrían aportar información relevante sobre el tema. Tras la lectura a texto completo de las referencias seleccionadas 6 se excluyeron del informe. Cuatro de las seis referencias (Kelly *et al.*, 2012; Caan *et al.*, 2014; Prat *et al.*, 2012 y Esserman *et al.*, 2012), se excluyeron al no realizar el análisis de la prueba del PAM50 con la tecnología *Nanostring Prosigna™* que es la evaluada en este informe. Las otras dos referencias excluidas, Zhao *et al.*, 2014 y Chia *et al.*, 2012, lo fueron por presentar el análisis estadístico de datos procedentes de poblaciones tanto ER+ como ER- con y sin tratamiento, con lo que no cumplía los criterios de inclusión, tanto por población como por agregación de datos.

Finalmente, el total de referencias incluidas en este apartado han sido 6. El esquema que muestra el flujo de selección de referencias se describe en la Figura 4.

Figura 4: Diagrama de flujo de la selección de estudios de Prosigna™



3.2.2 Descripción de los estudios y evaluación de la calidad metodológica

Las 6 referencias identificadas en la búsqueda bibliográfica recogían estudios que aportaban evidencia sobre la validez analítica y la validez clínica de la prueba. Los estudios localizados y su finalidad se recogen en la Tabla 10.

Tabla 10: Estudios primarios publicados sobre Prosigna™			
Referencia	Tipo de estudio	n	Finalidad
Nielsen <i>et al.</i> , 2014	Preclínico	43	Validación analítica de Prosigna™
Dowsett <i>et al.</i> , 2013 Sestak <i>et al.</i> , 2013	Cohortes Retrospectivo procedentes de un ensayo clínico aleatorizado (TransATAC)	1007	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+ (10 años) tratadas con terapia hormonal adyuvante
Gnant <i>et al.</i> , 2014 Filipits <i>et al.</i> , 2014	Cohortes Retrospectivo procedentes de un ensayo clínico aleatorizado (ABCSG-8)	1478	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+ (10 años) tratadas con terapia hormonal adyuvante
Sestak <i>et al.</i> , 2014	Cohortes Retrospectivo procedentes de dos ensayos clínicos aleatorizados (TransATAC + ABCSG-8)	2137	Validación clínica (carácter pronóstico) en pacientes postmenopáusicas ER+ (5 años) tratadas con terapia hormonal adyuvante

Todos los estudios localizados sobre de Prosigna™ han sido estudios pre-clínicos o retrospectivos, no identificándose en la búsqueda ningún estudio de carácter prospectivo. Los artículos de Dowsett *et al.*, 2013 y Sestak *et al.*, 2013, por una parte y los de Gnant *et al.*, 2014 y Filipits *et al.*, 2014, por otra, corresponden a los mismos estudios pero presentaban análisis de resultados complementarios por lo que no se han considerado duplicación de información y se han incluido en el informe.

Estudios de validez analítica de Prosigna™

Solamente se ha identificado un estudio sobre la validez analítica de la prueba (Nielsen *et al.*, 2014). Este estudio se ha centrado en la validación de la prueba genómica PAM50 usando el *nCounter Analysis System*, con la finalidad de poder realizar la prueba de manera descentralizada, en laboratorios locales de patología molecular, de manera simple y rutinaria. La validación se ha realizado valorando la reproducibilidad del ensayo en diferentes laboratorios y la precisión del mismo.

El diseño del estudio para la determinación de la precisión, consistió en la realización de la prueba a cinco muestras, cada una de ellas forma-

da por una mezcla de ARN procedentes de tumores de un mismo subtipo molecular. En concreto, en las cinco muestras, 3 eran del subtipo luminal (1 luminal A y 2 luminal B) y las dos restantes de los otros subtipos, HER2 y basal. Estas 5 muestras fueron divididas y analizadas en tres laboratorios diferentes: *British Columbia Cancer Agency (Vancouver)*, *Washington University (St. Louis)* y el laboratorio de *NanoString Technologies (Seattle)*. Para la estimación del índice ROR, al ser las muestras una mezcla de ARN de distintos tumores, se consideró como tamaño del tumor $\leq 2\text{cm}$ y el estado ganglionar de no afectación. Las muestras se analizaron en diferentes situaciones para medir la variabilidad debida a factores que pudieran afectar a su realización y precisión (repetibilidad), en concreto valoraron: el laboratorio, el operador, el lote de reactivo y cada uno de los ensayos individuales realizados. Cada una de las 5 muestras se analizó por duplicado en cada ensayo. En resumen, cada laboratorio completó 18 ensayos (9 por cada operador). Los resultados se incorporaron de forma cegada a una base de datos, que fue la utilizada en el análisis estadístico.

El diseño del estudio de reproducibilidad, también consideró las variables laboratorio, operador, el lote de reactivo y cada uno de los ensayos individuales realizados, aunque esta vez se utilizaron muestras procedentes de 43 tumores parafinados procedentes de pacientes ER+ diagnosticadas con carcinomas invasivos. De cada tumor se prepararon 3 grupos de secciones seriadas que fueron enviadas a cada laboratorio, cada grupo contenía una sección teñida con H-E de 4 μm y tres secciones de 10 μm . Cada laboratorio usó un lote diferente de reactivos para la extracción de ARN y todos usaron el mismo lote del *kit* de Prosigna™. Los resultados se incorporaron de forma cegada a una base de datos, que fue la utilizada en el análisis estadístico. Se consideró que la prueba era “altamente reproducible” si la desviación típica era $< 4,3$ unidades de ROR.

El estudio también, valoró la cantidad de ARNm en la muestra y las posibles interferencias producidas por la presencia de tejido no tumoral en la muestra, como parámetros a tener en cuenta a la hora de establecer la validez analítica de la prueba.

Valoración de la calidad metodológica: el estudio ha sido llevado a cabo de manera rigurosa con un buen diseño tanto en la validación de la reproducibilidad como en la validación de la precisión de la prueba, siguiendo los criterios estandarizados de verificación y validación de ensayos del *Clinical Laboratory and Standards Institute (CLSI)* para la evaluación de la precisión de dispositivos de diagnóstico *in vitro*. Se ha realizado en, al menos, 3 laboratorios diferentes, dos de ellos con certificación CLIA (*Clinical Laboratory Improvement Amendments*) que es la certificación necesaria para realizar pruebas diagnósticas con muestras humanas. El análisis estadístico de los datos ha sido exhaustivo y adecuado a los objetivos. El cegamiento a los

resultados de los otros centros y operarios, contribuiría al control del sesgo de realización.

Estudios de validez clínica de Prosigna™

La validez clínica de la prueba Prosigna™ se ha basado en estudios de cohortes de carácter retrospectivo que procedían de estudios clínicos aleatorizados. Estos estudios han analizado la prueba en las muestras tumorales disponibles, procedentes de pacientes incluidas en ensayos clínicos aleatorizados, realizados previamente y diseñados para comparar diferentes tratamientos de terapia hormonal. Las pacientes procedían básicamente de dos ensayos clínicos con los siguientes objetivos

Ensayo TransATAC: realizado en mujeres postmenopáusicas con cáncer de mama ER+ en estadios tempranos (procedente del ensayo clínico ATAC) de las cuales se disponía de muestras de tumores en parafina y estaban incluidas en uno de los brazos del ensayo con monoterapia. El ensayo ATAC tenía como objetivo comparar la eficacia y seguridad a 5 años del tratamiento con anastrozol, con tamoxifeno o con una combinación de ambos tratamientos.

Ensayo ABCSG-8: realizado en mujeres postmenopáusicas con cáncer de mama ER+ en estadios tempranos, con el objetivo de comparar si dos años de tratamiento con tamoxifeno seguidos de 3 años con anastrozol mejoraba los resultados de supervivencia frente a continuar esos 3 años con tamoxifeno.

De las 5 referencias identificadas, 2 han utilizado para la validación de la prueba la población procedente del estudio TransATAC (Dowsett *et al.*, 2013 y Sestak *et al.*, 2013), otras 2 han utilizado la población procedente del ABCSG-8 (Filipits *et al.*, 2014 y Gnant *et al.*, 2014), por último, la quinta referencia ha realizado un análisis estadístico uniando las dos poblaciones, TransATAC y ABCSG-8 (Sestak *et al.*, 2014).

Básicamente, los estudios de validación clínica han analizado la recurrencia a distancia de la enfermedad a 5 y 10 años. Para ello, han clasificado a las pacientes de estudio en dos grupos según el riesgo de recaída, grupo de bajo riesgo, de riesgo intermedio y de alto riesgo, y han estimado mediante análisis estadísticos de Kaplan-Meier la supervivencia o recurrencia a 5 y 10 años de cada grupo. Los estudios han comparado la capacidad pronóstica del índice ROR de Prosigna™ con otros índices pronóstico genómicos como Oncotype DX® o con índices pronósticos estimados a partir de factores de riesgo clínico-inmunohistopatológicos como los índices CTS (afectación ganglionar, tamaño de tumor, grado histopatológico, edad y tratamiento administrado), CLP (el tamaño y grado del tumor y la afectación ganglionar) e IHC4 (niveles de los marcadores ER, PgR, Ki67 y HER2). Esta comparación la han realizado mediante análisis estadísticos de índice de correlación

de Cox, “*c index*”, y mediante cambios en el “*Likelihood ratio*”, estos análisis han sido adecuados a la finalidad pretendida. Como en el caso anterior de la prueba EndoPredict®, se puede considerar que el diseño de los estudios de validez clínica era adecuado y pertinente al objetivo que se pretendía, con los puntos fuertes y debilidades ya descritos en el apartado correspondiente de la prueba EndoPredict®.

Una descripción detallada de cada uno de los estudios de validez clínica, así como, la valoración del riesgo de sesgo de los mismos se ha recogido en la Tabla 11.

En resumen, la evidencia de los principales estudios de valoración clínica de la prueba genómica Prosigna™ como factor pronóstico, siguiendo los criterios propuestos por SIGN, mostraba un nivel de evidencia 2+ (estudio de cohortes bien realizado con bajo riesgo de sesgos).

En cuanto a los posibles conflictos de interés, señalar que algunos de los autores de los artículos estaban vinculados a la empresa *NanoString Technologies, Inc.*, propietaria de la tecnología de estudio, o habían recibido algún tipo de financiación de dicha compañía. Los conflictos de interés han sido mencionados y descritos en todas las publicaciones

Tabla 11. Características de los estudios de validez clínica de Prosigna™ incluidos en el informe		
Estudio	Características del estudio	Diseño del estudio
Dowsett <i>et al.</i>, 2013 Sestak <i>et al.</i>, 2013 Financiación: <i>Breakthrough Breast Cancer, Royal Marsden National Institute for Health Research Biomedical Research Centre, Cancer Research UK Program Grant nº C569-10404 y NanoString Technologies, Inc.</i> Conflictos de interés: al menos 3 autores recibían honorarios de <i>NanoString Technologies, Inc.</i> Calidad metodológica: moderado/bajo riesgo de sesgo Nivel de evidencia: 2+	Población para la validación clínica de los índices: 1.017 mujeres postmenopáusicas con cáncer de mama ER+ tratadas con anastrozol o tamoxifeno y de las que se disponía muestras de ARN procedentes del ensayo clínico aleatorizado TransATAC En la población incluida había un 73,4 % de pacientes sin afectación ganglionar y un 26,6 % con afectación ganglionar. Un 77 % de los pacientes tenían un tamaño de tumor entre 1–3. Aproximadamente un 50 % de los pacientes (n = 513) habían sido tratados con anastrozol y otro 50 % (n = 494) con tamoxifeno. El 67,9 % había recibido radio-terapia Clasificación de los pacientes mediante dos índices pronóstico: ROR y CTS (índice clínico-patológico y de tratamiento: afectación ganglionar, tamaño de tumor, grado histopatológico, edad y tratamiento administrado) Medidas de resultado: tiempo libre de enfermedad desde la aleatorización hasta aparición de recurrencia a distancia	Diseño: estudio de cohortes retrospectivo procedentes de un ensayo clínico aleatorizado para la validación del índice de riesgo ROR Objetivo: validación del carácter pronóstico del índice ROR en mujeres postmenopáusicas con cáncer de mama ER+ Realización: se analizó la expresión génica y el índice ROR mediante la prueba Prosigna™ a una muestra de 1.007 ARN, de las 1.017 muestras disponibles, por presentar la calidad requerida por la prueba Prosigna™. Se comparó con otros índices pronóstico; en este informe sólo se recoge la información aportada por la prueba de estudio. La capacidad pronóstico se valoró mediante dos aproximaciones: • Cambios en el test estadístico LR (<i>Likelihood ratio</i>): para medir cuantitativamente la cantidad relativa de información de un índice en comparación con otro • Estimación del índice C (índice de concordancia): para medir la concordancia de tiempo hasta el evento, de manera que valores entre 0,5 y 1,0 indicaban incremento de predicción Se analizó la supervivencia mediante análisis Kaplan-Meier de los grupos de riesgo según el ROR Seguimiento: 10 años Limitaciones: Sestak <i>et al.</i>, 2013, solo han incluido los pacientes con información de los índices ROR, Oncotype DX y IHC4, el 84% de la población del ensayo TransATAC, que a su vez recoge una fracción de la población del ensayo ATAC. Podría existir una baja representación de tumores de pequeño tamaño, al necesitar suficiente tejido tumoral para la realización de todas las pruebas moleculares. El índice clínico-patológico CTS no es utilizado normalmente en la práctica clínica, aunque los factores de riesgo que incluye, sí lo son

Tabla 11. Características de los estudios de validez clínica de Prosigna™ incluidos en el informe (continuación)		
Estudio	Características del estudio	Diseño del estudio
Gnant <i>et al.</i> , 2014 Filipits <i>et al.</i> , 2014 Financiación: AstraZeneca y NanoString Technologies, Inc. Conflictos de interés: al menos 1 autor recibía honorarios de NanoString Technologies, Inc. Calidad metodológica: moderado/bajo riesgo de sesgo Nivel de evidencia: 2+	Población para la validación clínica: mujeres postmenopáusicas con cáncer de mama con receptores hormonales positivos (ER y/o PgR) en estadios tempranos procedentes del ensayo clínico aleatorizado ABCSG-8. El tamaño de la población fue de 1.478 y 1.246 para estimaciones a 10 y 15 años, respectivamente Los autores establecen un índice clínico-patológico (CLP) calculado con base en el tamaño y grado del tumor y la afectación o no de ganglios linfáticos, y estimado mediante una combinación lineal de los mismos Clasificación de los pacientes mediante dos índices pronóstico: ROR y CLP Medidas de resultado: supervivencia libre de enfermedad desde la aleatorización hasta aparición de recurrencia a distancia o muerte	Diseño: estudio de cohortes retrospectivo procedentes de un ensayo clínico aleatorizado para la validación del índice de riesgo ROR Objetivo: validación del carácter pronóstico del índice ROR en mujeres postmenopáusicas con cáncer de mama con receptores hormonales positivos Realización: se analizó la expresión génica y el índice ROR mediante la prueba Prosigna™ en una muestra de 1478 tumores para valorar el subtipo molecular y el índice ROR. También es estimó el índice pronóstico CLP ROR= 54,7690*(-0,0067 A+0,4317 B - 0,3172 C + 0,4894 D + 0,1981 E + 0,1133 F + 0,8926) - A = correlación Pearson para tipo basal - B = correlación Pearson para tipo HER2 - C = correlación Pearson para tipo luminal A - D = correlación Pearson para tipo luminal B - E = índice de proliferación - F = tamaño de tumor Mediante modelos estadísticos de Cox estimaron los HR y los intervalos de confianza al 95 % de los distintos factores de riesgo, y de los índices pronósticos; así como la cantidad relativa de información que aportaba cada índice, mediante cambios en el LR (<i>Likelihood ratio</i>). El estudio de Filipits <i>et al.</i>, 2014 mediante un análisis Kaplan-Meier ha estimado la probabilidad de supervivencia libre de enfermedad a 15 años. Los investigadores que estimaron el índice ROR y los subtipos moleculares estuvieron cegados, no teniendo acceso a los datos clínicos Seguimiento medio: 11 años Limitaciones: carácter retrospectivo. Se recoge una fracción de la población del ensayo ABCSG-8. Compararon el índice ROR con un índice clínico-patológico (CLP) que no es utilizado normalmente en la práctica clínica, aunque los factores de riesgo que incluye, si

Tabla 11. Características de los estudios de validez clínica de Prosigna™ incluidos en el informe (continuación)		
Estudio	Características del estudio	Diseño del estudio
Sestak <i>et al.</i>, 2014 Financiación: <i>Breakthrough Breast Cancer, Royal Marsden National Institute for Health Research Biomedical Research Centre, Cancer Research UK Program Grant nº C569-10404</i> Conflictos de interés: la mayor parte de los autores recibieron algún tipo de compensación de <i>NanoString Technologies, Inc.</i> Calidad metodológica: moderado/bajo riesgo de sesgo. Nivel de evidencia: 2+	Población para la validación clínica: 2.137 mujeres postmenopáusicas con cáncer de mama con receptores hormonales positivos en estadios tempranos procedentes de los ensayos clínicos aleatorizados ABCSG-8 y TransATAC y que no habían presentado recurrencia a distancia en los 5 primeros años de seguimiento Los autores establecen un índice clínico-patológico (CTS) calculado con base en el tamaño y grado del tumor, la afectación o no de ganglios linfáticos, edad y tratamiento administrado Clasificación de los pacientes mediante dos índices pronóstico: ROR y CTS Medidas de resultado: tiempo libre de enfermedad desde la aleatorización hasta aparición de recurrencia a distancia o muerte	Diseño: estudio de cohortes retrospectivo procedentes de dos ensayos clínicos aleatorizados para la validación del índice de riesgo ROR Objetivo: validación del carácter pronóstico del índice ROR en mujeres postmenopáusicas con cáncer de mama con receptores hormonales positivos a largo plazo (> 5 años de seguimiento) Realización: se analizó la expresión génica y el índice ROR mediante la prueba Prosigna™ en una muestra de 2.137 tumores parafinados disponibles. Se estimó el índice ROR y el índice CTS Se excluyeron los pacientes que habían tenido recurrencia a distancia durante los primeros 5 años de seguimiento. Mediante modelos estadísticos de Cox se valoró la asociación entre el índice ROR y la recurrencia a distancia, para lo que se estimó el HR (IC 95 %) Las curvas de supervivencia se estimaron mediante el análisis estadístico de Kaplan-Meier, y las curvas se truncaron a los 10 años debido a diferencias en los seguimientos de los dos ensayos clínicos incluidos en el estudio Mediante análisis de LR (<i>Likelihood ratio</i>) determinaron de manera cuantitativa la aportación relativa de información pronóstica del índice ROR en comparación con el índice CTS Limitaciones: carácter retrospectivo. Se han combinado cohortes procedentes de dos ensayos clínicos independientes. En el estudio TransATAC había tumores de mayor tamaño, de grado 3 y un mayor número de pacientes con más de 4 ganglios afectados. Por lo tanto, las características iniciales de las dos poblaciones no fueron las mismas (posible sesgo de selección)

3.2.3. Resultados clínicos de Prosigna™

Validez analítica

La información disponible sobre validez analítica procede de un único estudio, Nielsen *et al.*, 2014, los datos aportados en el estudio son los siguientes:

Precisión de la prueba

En general, todos los ensayos realizados con las 5 muestras de ARN de una mezcla de distintos tumores de un mismo subtipo molecular, mostraron que la prueba tenía una desviación típica inferior a 1 unidad de ROR (escala de 0 -100) y que más del 60 % de la variabilidad observada era intrínseca al ensayo, contribuyendo tanto el laboratorio como el operador a menos del 2 % de la variabilidad de la medida. En concreto, la variabilidad aportada por el laboratorio era inferior a 0,5 unidades de ROR, lo que se considera prácticamente despreciable. La distribución de los índices ROR determinados en los 108 ensayos independientes mostró que los valores se encontraban en un rango ≤ 4 unidades para cada una de las cinco muestras.

Se observó igualmente, una concordancia del 100 % entre los resultados esperados y los observados en cuanto al subtipo molecular de la muestra, así como en la clasificación al grupo de riesgo correspondiente.

Reproducibilidad de la prueba

En los tres laboratorios del estudio, se logró aislar ARNm en el 95 % (41/43), 93 % (40/43) y 100 % (43/43) de las 43 muestras de tejido tumoral analizadas. En los tres se logró obtener ARN de 40 de las muestras.

El factor de mayor aportación a la varianza de los resultados de las 41 muestras finalmente incluidas en el estudio por obtener ARN en más de un laboratorio, fue la “sección de tejido” que implicaba > 90 % de la varianza.

La concordancia, en cuanto al subtipo molecular y al grupo de riesgo, entre los tres laboratorios fue ≥ 90 %; no encontrándose ningún caso clasificado de bajo riesgo en un laboratorio y alto riesgo en otro, o viceversa, en los casos sin afectación ganglionar. Encontraron diferencias en la identificación de los subtipos moleculares en 2 muestras de riesgo intermedio/alto.

La comparación de la expresión génica de las réplicas de tejido en cada laboratorio, mostró una alta correlación con un coeficiente de Pearson de 0,98 (IC95 % 0,98-0,98). El estudio mostró asimismo, que las muestras de ARN que estaban dentro de las especificaciones de concentración de ARN necesaria, 62,5 ng-625 ng, pasaban la prueba.

Por último, ha mostrado que es importante la macrodissección para reducir las interferencias introducidas por el tejido no tumoral de las muestras. Esta interferencia podía dar lugar a una infraestimación del índice ROR o incluso un cambio en la determinación del subtipo molecular.

Validez analítica

Validación con la cohorte del ensayo Trans ATAC

Los estudios de Dowsett *et al.*, 2013 y Sestak *et al.*, 2013, realizados con la cohorte de pacientes procedentes del ensayo clínico TransATAC en la que se observaron 160 eventos de recurrencia a distancia.

Los análisis estadísticos de los datos mostraron que el índice ROR añadía información pronóstica a los índices clínico-patológicos. El índice ROR aportó, de manera significativa, un aumento en la información pronóstica sobre la recurrencia a distancia en la población de estudio frente al índice clínico-patológico, e incluso esa aportación pronóstica se observó en todos los subgrupos de pacientes considerados:

- pacientes sin afectación ganglionar
- pacientes con afectación ganglionar
- pacientes HER2-negativo
- pacientes HER2-negativo sin afectación ganglionar

Las comparaciones mediante el análisis de *Likelihood ratio* para comprobar la información pronóstica que el índice ROR aportaba al índice clínico-patológico CTS, se observó en la población global un LR - $\Delta\chi^2 = 33,9$, $p < 0,00$, que indicaba que el índice ROR aportaba un incremento de información pronóstica a la del índice CTS. El valor pronóstico encontrado en los pacientes HER2-negativos sin afectación ganglionar fue similar al encontrado en el conjunto de pacientes sin afectación ganglionar (un cambio en el LR - $\Delta\chi^2 = 23,4$ v $24,6$; $p < 0,001$, respectivamente).

Se observó una relación entre el valor del ROR y el riesgo de recurrencia a distancia, a mayor valor de ROR mayor riesgo. El índice ROR mostró una interpretación distinta para pacientes sin afectación ganglionar y pacientes con ganglios afectados; así, para ganglios negativos un riesgo de recurrencia del 10 % era pronosticado por un ROR de 42 mientras que este mismo riesgo en pacientes con ganglios positivos es pronosticado por un ROR de 25.

El análisis genético del PAM50 de la prueba Prosigna™ mostró que la distribución de subtipos moleculares en la población de estudio fue la siguiente:

- Pacientes sin afectación ganglionar: 529 luminal A (71,6 %) y 176 luminal B (23,8 %).
- Pacientes con 1-3 ganglios afectados: 132 luminal A (61,7 %) y 69 luminal B (32,2 %).
- Pacientes con ≥ 4 ganglios afectados: 31 luminal A (57,4 %) y 20 luminal B (37,0 %).

Se observó una tendencia de disminución del subtipo luminal A conforme aumentaba el número de ganglios afectados. La tasa de recurrencia a distancia fue mayor en pacientes luminal B que en luminal A (Tabla 12). En pacientes sin ganglios afectados, el subtipo luminal B mostró 4,78 veces más riesgo de recurrencia que el luminal A (HR 4,78; IC95 % 2,97-7,70, $p < 0,001$).

Tabla 12. Riesgo de recurrencia a distancia a 10 años en pacientes luminal A y B, con y sin afectación ganglionar	
Subtipo / afectación ganglionar	HR (IC95 %)
Luminal A / N0	-
Luminal B / N0	4,78 (2,97 – 7,70)
Luminal A / N 1–3	referencia
Luminal B / N 1–3	1,98 (1,10 – 3,59)
Luminal A / N > 4	2,90 (1,47 – 5,73)
Luminal B / N > 4	10,37 (5,42 – 19,84)

El análisis estadístico realizado por Sestak *et al.*, (2013) de la asociación entre los marcadores clínico-patológicos, el índice ROR y la recurrencia a distancia en función del tiempo, mostró que tanto entre los 0 – 5 años y los 5 – 10 años, el tamaño de tumor y la afectación ganglionar fueron las variables con mayor valor predictivo de la recurrencia a distancia, al menos hasta los 8 años de seguimiento. En cuanto a los índices de riesgo analizados, el ROR y los marcadores inmunohistoquímicos analizados en el estudio (ER, PgR, HER2 y Ki67) aportaban información pronóstica; sin embargo, el índice ROR era el índice con mayor valor pronóstico a largo plazo (entre los 5 y los 10 años de seguimiento) tanto en la población de estudio como en los distintos subgrupos, al discriminar mejor los pacientes de bajo riesgo y de alto riesgo de recurrencia a distancia.

Validación con la cohorte del ensayo ABCSG-8

Los estudios de Gnant *et al.*, 2014 y Filipits *et al.*, 2014, recogieron en los 11 años de seguimiento medio de la población un total de 155 eventos de recurrencia a distancia, 87 de los cuales ocurrieron entre los años 0–5 y los restantes 69 aparecieron a partir del quinto año de seguimiento.

Los modelos de regresión de Cox mostraron que los factores de riesgo con mayor poder predictivo fueron, el tamaño de tumor, el grado de tumor y la situación ganglionar; por lo que fueron los elegidos en el cálculo del índice CLP, utilizado como comparador clínico-patológico.

El *c-index*, calculado para valorar la correlación entre la supervivencia libre de recurrencia a distancia a 10 años pronosticada por el índice ROR (0,72; IC 95 % 0,68 – 0,76) y la observada, fue superior al del índice CLP (0,691; IC 95 % 0,646 – 0,737) lo que indicaba que aportaba mayor información que el índice basado solo en factores de riesgo clínico-patológicos. Resultados similares obtuvieron en el análisis de *Likelihood ratio* que mostró un cambio en el LR - $\Delta\chi^2 = 53,49$, $p < 0,0001$ a 10 años y un cambio en el LR - $\Delta\chi^2 = 15,23$, $p < 0,0001$ a los 15 años. Por subgrupos, se observó que el índice ROR aportaba mayor información pronóstica en pacientes con afectación ganglionar.

Las curvas de supervivencia mostraron una buena capacidad discriminatoria del índice ROR en los 3 grupos de riesgo establecidos a 10 y 15 años en la población global (Tabla 13). Resultados similares se observaron en la supervivencia de los tres grupos de riesgo por subgrupos de pacientes, con y sin afectación ganglionar (Tabla 14).

Tabla 13. Supervivencia libre de recurrencia a distancia en los tres grupos de riesgo en la cohorte de ABCSG-8 según el índice ROR		
Grupo de riesgo	Supervivencia % (IC95 %) 10 años	Supervivencia % (IC95 %) 15 años
Riesgo bajo	96,7 % (94,6 – 98,0)	97,6 % (94,7 – 98,9)
Riesgo intermedio	91,3 % (88, 1 – 93,8)	90,9 % (85,9 – 94,3)
Riesgo alto	79,9 % (75,7 – 83,4)	82,5 % (74,8 – 88,1)

Tabla 14. Supervivencia libre de recurrencia a distancia en los tres grupos de riesgo en pacientes sin afectación y con afectación ganglionar de la cohorte de ABCSG-8 según el índice ROR a 10 años de seguimiento		
Grupo de riesgo	Sin afectación ganglionar	Con afectación ganglionar
Riesgo bajo	96,6 % (94,4 – 97,9)	100 % (-)*
Riesgo intermedio	90,4 % (86,3 – 93,3)	93,6 % (86,9 – 97,0)
Riesgo alto	84,3 % (78,4 – 88,6)	76,1 % (69,9 – 81,2)
* no hubo ningún evento en este grupo (n=15)		

El índice ROR clasificó a los pacientes sin afectación ganglionar en los siguientes grupos: a un 47 % de ellos en riesgo bajo, a un 32 % en riesgo intermedio y a un 21 % en el grupo de alto riesgo. Por el contrario, clasificó al 67 % de los pacientes con afectación ganglionar en el grupo de alto riesgo y solamente a un 3 % en el grupo de bajo riesgo.

Los análisis de Kaplan-Meier a 15 años mostraron una diferencia significativa en la supervivencia libre de recurrencia entre los grupos de riesgo

establecidos por el índice ROR, entre los grupos de alto y bajo riesgo se encontró una HR = 6,90 ($p < 0,001$) y entre los grupos de intermedio y bajo riesgo la HR fue de 3,74 ($p = 0,002$).

La clasificación de subtipos establecida por el índice ROR mostró que a 10 y 15 años la supervivencia libre de recurrencia a distancia del subtipo luminal A era mayor que la del subtipo luminal B (Tabla 15), siendo esta diferencia estadísticamente significativa (HR = 2,85, IC 95 % 2,04 – 4,00 y HR = 2,14, IC 95 % 1,30 – 3,50 a 10 años y 15 años respectivamente, $p < 0,001$). Los pacientes luminal A presentaron menores índices ROR que los luminal B.

Tabla 15. Supervivencia libre de recurrencia a distancia por subtipos moleculares en la cohorte de ABCSG-8 según el índice ROR		
Subtipo molecular	Supervivencia % (IC95 %) 10 años	Supervivencia % (IC95 %) 15 años
Luminal A	93,9 % (IC 95 % 92,0 – 95,3)	92,8 % (89,6 – 95,1)
Luminal B	82,2 % (IC 95 % 77,8 – 85,8)	86,2 % (78,2 – 91,3)

Validación con la cohorte conjunta de los ensayos TransATAC y ABCSG-8

En el estudio se excluyeron aquellas pacientes que habían presentado recurrencia a distancia en los primeros 5 años de seguimiento, el ROR medio de las pacientes excluidas del estudio (con recurrencia en los 5 primeros años) fue significativamente superior al de las pacientes que no habían sufrido recurrencia y que estaban incluidas en este estudio (53,57 SD 20,4 vs 41,89 SD 19,5; $p < 0,001$, respectivamente). El índice ROR clasificó un 55,4 % de los pacientes en el grupo de bajo riesgo de recurrencia, un 25,2 % en el grupo de riesgo intermedio y un 19,5 % en el grupo de alto riesgo.

La tasa de recurrencia a distancia entre los años 5 y 10 de seguimiento, fue significativamente superior en la cohorte TransATAC que en la ABCGS-8, recogándose un total de 148 recurrencias a distancia a partir del quinto año de seguimiento.

En la población global, el índice CTS aportaba mayor cantidad de información pronóstica de recurrencia a partir del quinto año de seguimiento con una LR - $\Delta\chi^2 = 94,12$, el ROR también aportaba información pronóstica en este periodo de tiempo, aunque esa información era menor LR - $\Delta\chi^2 = 67,94$. Por el contrario, en pacientes sin afectación ganglionar el índice ROR aportaba más información pronóstica que el índice CTS, tanto en los análisis univariantes (LR - $\Delta\chi^2 = 30,95$ v 21,48) como bivariantes (LR - $\Delta\chi^2 = 17,25$ v 7,79). En las pacientes con afectación ganglionar el índice ROR aportaba menos información pronóstica, aunque esta aportación era significativa.

Las curvas de Kaplan-Meier mostraron que entre los años 5 y 10 de seguimiento, las pacientes del grupo de alto riesgo y de riesgo intermedio

presentaban, respectivamente, un riesgo de recurrencia 6,9 y 3,3 veces mayor que las pacientes clasificadas en el grupo de bajo riesgo. Un riesgo similar se observó en el subgrupo de pacientes sin afectación ganglionar y HER2-negativo (Tabla 16).

Tabla 16. Riesgo de recurrencia a distancia por grupos de riesgo en la cohorte mixta TransATAC/ABCSG-8 según el índice ROR a 5 y 10 años de seguimiento			
Grupo de riesgo	Probabilidad de recurrencia (%; IC95 %)	Población global (HR, IC95 %)	Subgrupo N0, HER2- (HR, IC95 %)
Riesgo bajo	2,0 % (1,3 – 3,2)	referencia	referencia
Riesgo intermedio	9,0 % (6,3 – 13,0)	3,26 (2,07 – 5,13)	3,75 (2,19 – 6,41)
Riesgo alto	11,5 % (6,8 – 19,0)	6,9 (4,54 – 10,47)	5,49 (2,92 – 10,35)
N0: sin afectación ganglionar			

Utilidad clínica

No se ha identificado ningún estudio con el diseño adecuado para poder valorar la utilidad de la prueba Prosigna™, ya que para validar o establecer la utilidad clínica se necesitarían estudios que valoraran de manera prospectiva la recurrencia de la enfermedad teniendo en consideración el índice ROR en la decisión terapéutica.

Impacto en la decisión clínica

No se han identificado estudios que valoraran el impacto de la realización de la prueba Prosigna™ en la decisión terapéutica.

4. Discusión

Validez analítica

Actualmente no existe una prueba referente (“*gold estándar*”) para estimar la expresión génica. Esta ausencia hace que no haya disponibles estimaciones de ratios de falsos positivos o falsos negativos. Las pruebas que actualmente se usan para esta estimación son la hibridación (*microarrays*) y la amplificación del ARN mediante PCR utilizando una transcriptasa reversa (RT-PCR). Existen factores que afectan a la reproducibilidad y precisión de las medidas de expresión génica. Entre ellas se encuentra la toma de muestras, ya que al extraer tejido tumoral puede arrastrar consigo tejido sano con lo que diluye los ARNm que se expresen de forma específica o mayoritaria en las células tumorales, y la necesidad de un número mínimo de células tumorales para obtener cantidad suficiente ARNm para la realización de las pruebas.

Otra fuente de variabilidad es la preparación y almacenaje de las muestras. El ARN se degrada fácilmente por lo que la calidad y los lotes de los productos utilizados durante su manipulación pueden afectar al proceso de extracción. El ARN se degrada por cambios en el pH y por la presencia de ribonucleasas, enzimas presentes en las células y que a veces también se encuentran en los reactivos químicos utilizados en la extracción del ARN. Puesto que la validez analítica es difícil de evaluar, los estudios se centran en demostrar:

- la reproducibilidad de la prueba (mismo test en un laboratorio varias veces con la misma muestra de tejido o con diferentes muestras y el mismo test en diferentes laboratorios).
- la reproducibilidad a lo largo del tiempo.
- los factores que afectan a la hora de realizar el test.

En el caso de EndoPredict®, los estudios realizados para demostrar la reproducibilidad de la muestra han ido encaminados a mostrar que siguiendo un protocolo estandarizado, la prueba puede realizarse en diferentes laboratorios manteniendo unos resultados similares que no afectan a la clasificación correcta de los pacientes (en cuanto al riesgo de recurrencia a distancia de la enfermedad), aunque dicha reproducibilidad, está basada en un único estudio realizado en 7 laboratorios y tal y como los propios autores mencionan, estos resultados deben de ampliarse. Los autores encontraron que la diferencia en los valores del EP entre los análisis locales y el de referencia eran $\leq 0,5$ unidades, por lo que en aquellas muestras donde los índices EP estimados estén dentro de 0,5 unidades cerca del punto de corte (valor 5), es decir, entre 4,5 y 5,5 podrían no clasificar correctamente el riesgo de los pacientes.

EndoPredict® ha sido analizado tanto en biopsias como en piezas quirúrgicas, mostrando que el método de obtención no afectaba al resultado. Este hecho favorecería la realización de la prueba antes de la cirugía sabiendo con antelación, el riesgo de la paciente y por consiguiente, la terapia adyuvante más adecuada.

Prosigna™, al igual que EndoPredict®, ha sido validada analíticamente, determinando su reproducibilidad y precisión. La prueba ha mostrado una buena concordancia, tanto entre los valores de la prueba obtenidos en cada uno de los laboratorios implicados en dicha validación y los valores de referencia, así como en la clasificación de los tumores estudiados en todos ellos.

Es importante señalar que, en los estudios de validez analítica los laboratorios implicados, son laboratorios de patología molecular, es decir, laboratorios con un cierto grado de especialización y experiencia en este tipo de técnicas de biología molecular. La representatividad de la muestra de laboratorios para asumir la validez externa del procedimiento podría no ser la más adecuada, es decir, habría que analizar si estos laboratorios son iguales a los disponibles en nuestro sistema sanitario.

La valoración cuantitativa de ARN no es un proceso fácil ni sencillo por lo que para una buena realización descentralizada de la prueba EndoPredict® es necesario disponer de un sistema de PCR cuantitativa estandarizada además de un protocolo reproducible de aislamiento de ARN a partir de muestras parafinadas. En el caso de Prosigna™ es necesaria la adquisición del sistema *nCounter Analysis System con sus reactivos* y seguir el protocolo establecido.

Hay que reseñar, que un paso importante en la estimación de estos índices parece ser la cantidad de tejido no tumoral presente en la sección analizada, ya que implican una contaminación en la muestra de ARN que podría suponer una fuente de sesgo en este tipo de factores pronóstico. Parece ser, que el impacto de esta contaminación con tejido sano varía o depende de los genes incluidos en la prueba genómica, lo que requeriría la implantación de estrategias de control de calidad que garanticen un contenido mínimo de ARN procedente de células cancerosas (Elloumi *et al.*, 2011).

Las dos pruebas utilizan tecnología diferente para la medida de la expresión génica; la prueba Prosigna™ presenta una medida desde el punto de vista técnico, con menos manipulación, ya que la se cuantifica la expresión en solución una vez purificado el ARNm, sin necesidad de sintetizar ADNc ni amplificar por PCR. Esta reducción de pasos a la hora de medir la expresión reduce por consiguiente la probabilidad de degradación del ARN aunque requiere mayor inversión en instrumentación analítica.

Uno de los atractivos de esta segunda generación de pruebas genómicas es que están diseñadas para que sean realizadas de forma local en los laboratorios interesados, lo que en principio supondría un abaratamiento de

los costes de la prueba, al evitar el envío a laboratorios de referencia. Si bien, la validación de su descentralización supone la adquisición de los instrumentos y reactivos determinados por el comerciante, implicando una inversión inicial considerable, y la adquisición de reactivos y el mantenimiento de los instrumentos.

Validez clínica

A diferencia de las plataformas genómicas de primera generación, las incluidas en este informe han sido validadas en estudios que muestran un mayor control de sesgos, lo que genera una evidencia científica con mayor calidad metodológica.

Las dos pruebas genómicas han sido validadas clínicamente en estudios retrospectivos utilizando la población procedente de ensayos clínicos aleatorizados previos que se realizaron con objetivos terapéuticos. Este tipo de diseño ha sido propuesto como estudio "prospectivo-retrospectivo" por Simon *et al.*, 2009, y a pesar de ser de carácter retrospectivo, garantiza que el seguimiento de los pacientes se llevó a cabo de forma prospectiva favoreciendo la recogida de información y por lo tanto, disponiendo de registros adecuados.

La necesidad de establecer este nuevo diseño se justificaría por el hecho de que la validación mediante ensayos clínicos aleatorizados o cohortes prospectivas de marcadores moleculares para enfermedades como el cáncer, implicaría un largo tiempo de seguimiento (5-10 años), lo que supondría un alto coste económico y el riesgo de que la tecnología en ese tiempo se convierta en obsoleta. Estos aspectos son importantes y deben ser considerados con el fin de llevar a cabo los estudios más adecuados y realizables, sin obviar los riesgos de sesgos implícitos al diseño, y planteando desde el principio, el análisis de las posibles variables confusoras para su control.

Las pruebas genómicas de carácter pronóstico o predictivo en el cáncer de mama sobretodo en el cáncer ER+ han mostrado que este valor pronóstico/predictivo está limitado a los primeros 5 años de seguimiento, disminuyendo o perdiendo dicha capacidad con el tiempo. Esta dependencia del tiempo hace que sea necesario disponer de pruebas y herramientas con carácter pronóstico a más largo plazo, ya que el riesgo de recurrencia a distancia se extiende más allá de los 20 años (Esserman *et al.*, 2011). Esta necesidad podría justificar la aparición de pruebas genómicas de segunda generación que han incorporando las variables clínicas con el fin de incrementar el valor pronóstico de las pruebas y que éste sea a más largo plazo (Zhao *et al.*, 2014). Las pruebas genómicas, Oncotype DX® y MammaPrint®, han mostrado buena capacidad pronóstica/predictiva a 5 años de seguimiento disminuyendo esta capacidad a mayor seguimiento. Por el contrario, la incorporación de parámetros clínico-patológicos parece mejorar la capaci-

dad pronóstica de las plataformas genómicas (Zhao *et al.*, 2014; Sestak *et al.*, 2013). Por lo tanto, la estimación a largo plazo del riesgo de recurrencia a distancia es importante al encontrarse actualmente, en discusión y estudio, tanto la identificación de la población femenina que se vería beneficiada de una extensión del tratamiento de terapia hormonal, como la cantidad de tiempo a extender más allá de los 5 años establecidos (Knauer, Filipits & Dubsky, 2014).

Actualmente, se están llevando a cabo estudios de concordancia entre las pruebas genómicas disponibles, con la finalidad de comparar entre ellas su capacidad pronóstica/predictiva, pese a ser evidente el poco solapamiento entre los genes incluidos en las mismas (Huang *et al.*, 2014). La comparación de EndoPredict® con el Índice de Recurrencia (IR) de Oncotype DX® (Varga *et al.*, 2013) mostró que existían una moderada/baja correlación entre ambas pruebas, aunque es estadísticamente significativa ($p < 0,01$), con un coeficiente de Pearson entre IR y EP de 0,65, y entre IR y EPclin de 0,45. Esta baja correlación se estimó en una población de 34 mujeres con diferente afectación ganglionar, grado y tamaño de tumor. Además, encontraron una concordancia en la clasificación de las dos pruebas del 76% (26 de los 34 casos) cuando se agrupaban en el mismo grupo (alto-intermedio) los casos clasificados de riesgo intermedio por el IR. La comparación entre PAM50, Oncotype DX®, MammaPrint® y EndoPredict® realizada en una población general (n = 912) obtenida de seis bases de datos de *microarrays*, mostró que todas las pruebas discriminaban de manera significativa entre los grupos de riesgo que establecían en relación a la supervivencia libre de recurrencia, y que el PAM50 mostró el índice *c* más alto. Considerando, tanto para PAM50 como Endopredict®, solo la parte correspondiente a la expresión génica (Tabla 17).

Tabla 17. Comparación de la capacidad pronóstica de varias pruebas genómicas (c-index).		
Prueba	Índice c	IC 95%
PAM50 (RORs)	0,658	0,64 – 0,68
EndoPredict® (EP)	0,648	0,63 – 0,67
Oncotype DX®	0,648	0,63 – 0,67
MammaPrint®	0,612	0,60 – 0,63

La prueba EndoPredict® validada en tres cohortes diferentes, y la prueba Prosigna™, validada en dos cohortes, han mostrado que en todas ellas era una factor pronóstico independiente. En la población del ensayo GEICAM/9906,

tratada con quimioterapia adyuvante, la prueba EndoPredict® ha presentado mayor rendimiento en el grupo de mujeres premenopáusicas con afectación ganglionar. La pregunta que surge en este caso, es si la prueba en esta población aportaría o implicaría una modificación en el seguimiento de las pacientes, ya que se trata de mujeres que ya están tratadas con quimioterapia y consideradas de alto riesgo. La prueba en principio iría destinada principalmente a la identificación de aquellas pacientes que tuvieran un riesgo bajo de recurrencia a distancia y por consiguiente, evitaran la quimioterapia y sus efectos adversos, presentando así, mayor utilidad la prueba.

La prueba Prosigna™, parte de la evidencia del valor pronóstico de la expresión de los 50 genes incluidos en PAM50 (Ward *et al.*, 2013), y supone una adaptación de una prueba previa a una nueva tecnología el *nCounter Analysis System*, que no requiere ni la utilización de retrotranscriptasa, ni de PCR, utilizando los índices ROR propuestos para la estimación del riesgo según el PAM50 determinado previamente mediante RT-PCR (Parker *et al.*, 2009). En los estudios de valoración clínica del índice ROR ha mostrado que aumenta la cantidad de información pronóstico al combinarlo con la información procedente de los índices clínico-patológicos CTS y CLP. Se observa una diferencia en la capacidad de predicción entre los dos estudios, así el incremento a 10 años en el estudio realizado en la cohorte de TransATAC es de $LR - \Delta\chi^2 = 33,9$ frente al $LR - \Delta\chi^2 = 53,49$ encontrado en la cohorte de ABCSG-8. Esta diferencia en la capacidad pronóstica podría deberse al hecho de que los índices clínico-patológicos utilizados en ambos estudios no son iguales; así, en el primer caso el índice utilizado es el CTS que incluye los mismos factores de riesgo que el CLP, utilizado en el segundo estudio, pero además incluye factores de riesgo como la edad y el tratamiento administrado. En consecuencia, podría pensarse que el índice clínico-patológico CTS presentaría mayor capacidad pronóstica que el CLP y por consiguiente, el índice ROR aportaría menor información pronóstico al CTS que al CLP.

En resumen, las dos pruebas genómicas valoradas en este informe, que combinan niveles de expresión génica con factores de riesgo clínico-patológicos, parecen aportar información pronóstica adicional, a la aportada por los factores clínico-patológicos convencionales, y a más largo plazo.

5. Conclusiones

Validez analítica

- Las pruebas genómicas EndoPredict® y Prosigna™ han presentado validez analítica en cuanto a precisión y reproducibilidad entre laboratorios.

Validez clínica

- Los índices EP y EPclin de EndoPredict® son factores pronóstico independientes que añaden información pronóstico a los índices clínico-patológicos, en mujeres pre y postmenopáusicas ER-positivo, HER2-negativo con y sin afectación ganglionar, a 10 años de seguimiento. La prueba presenta una buena capacidad discriminatoria entre los dos grupos de riesgo que establece: alto y bajo.
- El índice ROR de Prosigna™ es un factor pronóstico independiente que añade información pronóstico a los índices clínico-patológicos, en mujeres postmenopáusicas HR-positivo con y sin afectación ganglionar a 10 años de seguimiento. Y muestra una buena capacidad discriminatoria entre los tres grupos de riesgo que establece: bajo, intermedio y alto.
- La realización de la prueba EndoPredict® parece tener impacto en la decisión terapéutica, reduciendo el número de pacientes tratados con quimioterapia adyuvante.

Utilidad clínica

- No se dispone de evidencia sobre la utilidad clínica de ninguna de las dos pruebas

6. Referencias

- Bertucci, F., Finetti, P., Viens, P. & Birnbaum, D. (2014). *Cancer Letters*, 355, 70-75. doi: 10.1016/j.canlet.2014.09.014.
- Caan, B.J., Sweeney, C., Habel, L.A., Kwan, M.L., Kroenke, C.H., Weltzien, E.K., ...Bernard, P.S. (2014). Intrinsic subtypes from the PAM50 gene expression assay in a population-based breast cancer survivor cohort: prognostication of short- and long-term outcomes. *Cancer Epidemiology, Biomarkers & Prevention*, 23, 725-34. doi: 10.1158/1055-9965.EPI-13-1017.
- Chia, S.K., Bramwell, V.H., Tu, D., Shepherd, L.E., Jiang, S., Vickery, T., ... Nielsen, T.Q. (2012). A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clinical Cancer Research*, 18, 4465-4472. doi: 10.1158/1078-0432.CCR-12-0286.
- Denkert, C., Kronenwett, R., Schlake, W., Bohmann, K., Penzel, R., Weber, K.E.,... Dietel, M. (2012). Decentral gene expression analysis for ER+/Her2- breast cancer: results of a proficiency testing program for the EndoPredict assay. *Virchows Archiv*, 460, 251-259. doi: 10.1007/s00428-012-1204-4.
- Dowsett, M., Sestak, I., Lopez-Knowles, E., Sidhu, K., Dunbier, A.K., Cowens, W., ... Cuzick, J. (2013). Comparison of PAM50 Risk of Recurrence Score with Oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology*, 31, 2783-2790. doi: 10.1200/JCO.2012.46.1558.
- Dubsky, P., Brase, J.C., Jakesz, R., Rudas, M., Singer, C.F., Greil, R., ... Filipits, M. (2013a). The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *British Journal of Cancer*, 109, 2959-2964. doi:10.1038/bjc.2013.671.
- Dubsky, P., Filipits, M., Jakesz, R., Rudas, M., Singer, C.F., Greil, R., ... Gnant, M. (2013b). EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Annals of Oncology*, 24, 640-647. doi:10.1093/annonc/mds334.
- Elloumi, F., Hu, Z., LI, Y., Parker, J.S., Gulley, M.L., Amos, K.D., Troester, M.A. (2011). Systematic bias in genomic classification due to contaminating non-neoplastic tissue in breast tumor samples. *BMC Medical Genomics*, 4, 54. doi:10.1186/1755-8794-4-54.
- Esserman, L.J., Berry, D.A., Cheang, M.C., Yau, C., Perou, C.M., Carey, L.,... Hylton, N. (2012). Chemotherapy response and recurrence-free survival in

neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Research and Treatment*, 132, 1049-1062. doi: 10.1007/s10549-011-1895-2.

Esserman, L.J., Moore, D.H., Tsing, P.J., Chu, P.W., Yau, C., Ozanne, U., ... Benz, C.C. (2011). Biologic markers determine both the risk and the timing of recurrence in breast cancer. *Breast Cancer Research and Treatment*, 129, 607-616. doi: 10.1007/s10549-011-1564-5.

Filipits, M., Nielsen, T.O., Rudas, M., Greil, R., Stöger, H., Jakesz, R., ... Gnant, M. (2014). The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clinical Cancer Research*, 20, 1298-1305. doi: 10.1158/1078-0432.CCR-13-1845.

Filipits, M., Rudas, M., Jakesz, R., Dubsy, P., Fitzal, F., Singer, C.F., ... Gnant, M. (2011). A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clinical Cancer Research*, 17, 6012-6020. doi: 10.1158/1078-0432.CCR-11-0926.

Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Panel members. (2013). Personalizing the treatment of women with early breast cancer: Highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2013. *Annals of Oncology*, 24, 2206-2223. doi: 10.1093/annonc/mdt303.

Gnant, M., Filipits, M., Greil, R., Stoeger, H., Rudas, M., Bago-Horvath, Z., ... Nielsen, T.O. (2014) Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 risk of Recurrence Score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Annals of Oncology*, 25, 339-345. doi: 10.1093/annonc/mdt494.

Hayden, J.A., Côté, P., Bombardier, C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine*, 144, 427-437.

Huang, C-C., Tu, S-H., Lien, H-H., Huang, C-S., Huang, C-J., Lai, L-c.,... Chuang, E.Y. (2014). Refinement of breast cancer risk prediction with concordant leading edge subsets from prognostic gene signatures. *Breast Cancer Research and Treatment*, 147, 353-370. doi: 10.1007/s10549-014-3104-6.

Kelly, C.M., Bernard, P.S., Krishnamurthy, S., Wang, B., Ebbert, M.T.W., Bastien, R.R.L., ... Puztai, L. (2012). Agreement in Risk Prediction between the

21-gene Recurrence Score Assay (Oncotype DX®) and the PAM50 Breast Cancer Intrinsic Classifier™ in early-stage estrogen receptor-positive breast cancer. *The Oncologist*, 17, 492-498. doi: 10.1634/theoncologist.2012-0007.

Knauer, M., Filipits, M., Dubsky, P. (2014). Late recurrences in early breast cancer: from whom and how long is endocrine therapy beneficial? *Breast Care*, 9, 97-100. doi: 10.1159/000362482.

Kronenwett, R., Bohmann, K., Prinzler, J., Sinn, B.V., Haufe, F., Roth, C.,... Denkert, C. (2012). Decentral gene expression analysis: analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test. *BMC Cancer*, 5, 456-. doi: 10.1186/1471-2407-12-456.

Martin, M., Brase, J.C., Calvo, L., Krappmann, K., Ruiz-Borrego, M., Fisch, K., ... Rodriguez-Lescure, A. (2014). Clinical validation of the EndoPredict test in node-positive chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM/9906 trial. *Breast Cancer Research*, 16, R38. doi: 10.1186/bcr3642.

Müller, B.M., Brase, J.C., Haufe, F., Weber, K.E., Budzies, J., Petry, C., ... Denkert, C. (2012). Comparison of the RNA-based EndoPredict multigene test between core biopsies and corresponding surgical breast cancer sections. *Journal of Clinical Pathology*, 65, 660-662. doi: 10.1136/jclinpath-2012-200716.

Müller, B.M., Keil, E., Lehmann, A., Winzer, K-J., Ritcher-Ehrenstein, C., Prinzler, J., ... Denkert, C. (2013). The EndoPredict gene-expression assay in clinical practice-performance and impact on clinical decisions. *PLoS ONE*, 8, e-68252. doi: 10.1371/journal.pone.0068252.

Nielsen, T., Wallden, B., Schaper, C., Ferree, S., Liu, S., Gao, D., ... Storhoff, J. (2014). Analytical validation of PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BioMed Central Cancer*, 14, 177. doi: 10.1186/1471-2407-14-177.

Poremba, C., Uhlenendorff, J., Pfitzner, B.M., Hennig, G., Bohmann, K., Bojar, H., ... Denkert, C. (2014). Preanalytical variables and performance of diagnostic RNA-based gene expression analysis in breast cancer. *Virchows Archiv*, 465, 409-417. doi: 10.1007/s00428-014-1652-0.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., ... Bernard, P.S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27, 1160-1167. doi: 10.1200/JCO.2008.18.1370.

Prat, A., Parker, J.S., Fan, C., Cheang, M.C., Miller, L.D., Bergh, J., ... Perou, C.M. (2012). Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Annals of Oncology*, 23, 2866-2873. doi: 10.1093/annonc/mds080.

Sanderson, S., Zimmern, R., Kroese, M., Higgins, J., Patch, C., Emery, J. (2005). How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. *Genetic in Medicine*, 7, 495-500. doi: 10.1097/01.gim.0000179941.44494.73.

Schnitt, S.J. (2010). Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Modern Pathology*, 23, S60-S64. doi: 10.1038/modpathol.2010.33.

Sestak, I., Cuzick, J., Dowsett, M., Lopez-Knowles, E., Filipits, M., Dubsky, P., ... Gnant, M. (2014). Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian Breast and Colorectal Cancer study Group 8 and Arimidex, Tamoxifen Alone or in Combination Randomized Trials Using the PAM50 Risk of Recurrence Score. *Journal of Clinical Oncology*, 32, doi: 10.1200/JCO.2014.55.6894.

Sestak, I., Dowsett, M., Zabalgo, L., Lopez-Knowles, E., Ferree, S., Cowens, J.W., Cuzick, J. (2013). Factors predicting late recurrence for estrogen receptor-positive breast cancer. *Journal of the National Cancer Institute*, 105, 1504-1511. doi: 10.1093/jnci/djt244.

Siegel, R., Ma, J., Zou, Z., Jemal, A. (2014). Cancer statistics. 2014. *CA: a Cancer Journal for Clinicians*, 64, 9-29. doi: 10.3322/caac.21208.

Simon, R.M., Paik, S., Hayes, D.F. (2009). Use of Archived Specimens in Evaluation of Prognostic and Predictive Biomarkers. *Journal of The National Cancer Institute*, 101, 1446 – 1452. doi: 10.1093/jnci/djp335.

Sociedad Española de Oncología Médica (SEOM). (2014). Las cifras del cáncer en España 2014. Madrid. Recuperado de: http://www.seom.org/seomcms/images/stories/recursos/Las_cifras_del_cancer_2014.pdf

Urrutia, G. & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metanálisis. *Medicina Clínica*, 135, 507-511.

Varga, Z., Sinn, P., Fritzsche, F., von Hochstetter, Noske, A., Schraml, P., ... Moch, H. (2013). Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. *Plos ONE*, 8, e58483. doi:10.1371/journal.pone.0058483.

Ward, S., Scope, A., Rafia, R., Pandor, A., Harnan, S., Evans, P., Wyld, L. (2013). Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technology Assessment* 2013; 17(44): 1-302.

Zhang, M.H., Man, H.T., Zhao, X.D., Dong, N., Ma, S.L. (2014). Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomedical Report*, 2, 41-52.

Zhao, X., Rødland, E.A., Sørli, T., Volla, H.K., Russnes, H.G., Kristensen, V.N., Lingjærde, O.C., Børresen-Dale, A.L. (2014). Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer*, 14, 211. doi: 10.1186/1471-2407-14-211.

7. Anexos

Anexo 1. Estrategias de búsquedas

EndoPredict®

Ovid MEDLINE(R) without Revisions <1996 to October Week 2 2014>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <October 17, 2014>, Ovid MEDLINE(R) Daily Update <October 17, 2014>

1. exp *Breast Neoplasms/
2. ((breast* or mamma*) and (cancer* or tumor?r? or neoplas* or malign* or carcinoma?)).ti,ab.
3. Endopredict*.af.
4. (1 or 2) and 3

Embase hasta octubre 2014

1. exp *Breast Neoplasms/
2. breast*:ab,ti OR mamma*:ab,ti AND (cancer*:ab,ti OR tumor\$:ab,ti OR tumour\$:ab,ti OR neoplas*:ab,ti OR malign*:ab,ti OR carcinoma\$:ab,ti)
3. 1 OR 2
4. endopredict:ab,ti
5. 3 AND 4
6. 3 AND 4 NOT [medline]/lim
7. 6 NOT ('conference abstract'/it OR 'editorial'/it)

Pubmed hasta Octubre 2014

((("prosigna"[Title/Abstract]))) OR (((("endopredict"[Title/Abstract] OR "endopredict assay"[Title/Abstract] OR "endopredict assays"[Title/Abstract])))

Prosigna™

Ovid MEDLINE(R) without Revisions <1996 to October Week 4 2014>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <October 30, 2014>, Ovid MEDLINE(R) Daily Update <October 30, 2014>

1. exp *Breast Neoplasms/
2. ((breast* or mamma*) adj3 (cancer* or tumor* or neoplas* or malign* or carcinoma?)).ti,ab.
3. 1 or 2
4. (“ROR” or (risk adj2 recurrence)).ti,ab.
5. (prosigna* or PAM?50).af.
6. 4 and 5
7. 3 and 6

Embase hasta octubre 2014

1. ‘breast tumor’/exp/mj
2. ((breast* OR mamma*) NEAR/3 (cancer* OR tumor\$ OR tumour\$ OR neoplas* OR malign* OR carcinoma\$)):ab,ti
3. 1 OR 2
4. ‘ror’:ab,ti OR (risk NEAR/2 recurrence):ab,ti
5. prosigna*:df,ab,dn,ti OR pam50:df,ab,dn,ti OR ‘pam-50’:df,ab,dn,ti OR (pam NEAR/2 50):df,ab,dn,ti
6. 4 AND 5
7. 3 AND 6
8. 7 NOT [medline]/lim
9. 8 AND ‘conference abstract’/it
10. 8 NOT ‘conference abstract’/it

Pubmed hasta Octubre 2014

Search (((“Breast Neoplasms”[Majr]) OR ((((((((((“carcinoma”[Title/Abstract]) OR “carcinomas”[Title/Abstract]) OR malign*[Title/Abstract]) OR neoplasm*[Title/Abstract]) OR “tumor”[Title/Abstract]) OR “tumour”[Title/Abstract]) OR “cancer”[Title/Abstract]) OR cancer*[Title/Abstract])) AND ((breast*[Title/Abstract]) OR mamma*[Title/Abstract])))) AND ((((((“risk”[Title/Abstract]) AND “recurrence”[Title/Abstract]) OR “ror”)) AND (((“prosigna”) OR “pam 50”) OR “pam50”)))

Anexo 2. Listado de comprobación de sesgos de estudios pronósticos

Checklist: NICE

Study identification Include author, title, reference, year of publication				
Guideline topic:		Review question no:		
Checklist completed by:				
		Circle or highlight one option for each question		
1.1	The study sample represents the population of interest with regard to key characteristics, sufficient to limit potential bias to the results	Yes	No	Unclear
1.2	Loss to follow-up is unrelated to key characteristics (that is, the study data adequately represent the sample), sufficient to limit potential bias	Yes	No	Unclear
1.3	The prognostic factor of interest is adequately measured in study participants, sufficient to limit potential bias	Yes	No	Unclear
1.4	The outcome of interest is adequately measured in study participants, sufficient to limit potential bias	Yes	No	Unclear
1.5	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest	Yes	No	Unclear
1.6	The statistical analysis is appropriate for the design of the study, limiting potential for the presentation of invalid results	Yes	No	Unclear

