AETSA — EVALUACIÓN DE TECNOLOGÍAS SANITARIAS DE ANDALUCÍA

Validez diagnóstica del algoritmo de *The Fetal Medicine Foundation*para el cribado de preeclampsia pretérmino durante el primer trimestre de embarazo

INFORME_1

2021



Validez diagnóstica del algoritmo de *The Fetal Medicine Foundation* para el cribado de preeclampsia pretérmino durante el primer trimestre de embarazo

Revisión sistemática y metaanálisis

AETSA – Evaluación de Tecnologías Sanitarias de Andalucía Secretaría Gral. de Investigación, Desarrollo e Innovación en Salud Avda. de la Innovación s/n. Edificio ARENA 1. 41020 Sevilla España – Spain https://www.aetsa.org aetsa.csalud@juntadeandalucia.es

Autora: Isabel Ma Martínez Férez

Documentación: Antonio Romero Tabares

Revisión externa: Este trabajo se ha beneficiado de forma importante de la revisión realizada por el Dr. Lucas Cerrillos González. Jefe de Sección de Obstetricia y Urgencias del Hospital de la Mujer (Hospital Universitario virgen del Rocío). Sevilla.

<abril de 2022 >



Martínez Férez, Isabel María

Validez diagnóstica del algoritmo de The Fetal Medicine Foundation para el cribado de preeclampsia pretérmino durante el primer trimestre de embarazo Revisión sistemática y metaanálisis. Isabel Mª Martínez Férez — Sevilla: Consejería de Salud y Familias.

89 p; Informes de Evaluación de Tecnologías Sanitarias.

1. Preeclampsia / prevención y control 2. Preeclampsia / diagnóstico 3. Complicaciones del embarazo 4. Diagnóstico / métodos I. Andalucía. Consejería de Salud y Familias.

Conflicto de interés:

La autora declara 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.

The Fetal Medicine Foundation es una organización benéfica registrada en el Reino Unido que tiene como objetivo mejorar la salud de las mujeres embarazadas y sus bebés a través de la investigación y la capacitación en medicina fetal.

The Fetal Medicine Foundation https://fetalmedicine.org/ 137 Harley Street London, W1G 6BG United Kingdom



Depósito legal: SE 1022-2022

DOI: http://doi.org/10.52766/SAGJ1958

Índice

Resumen ej	ecutivo 5
Abreviatura	s
Glosaario	9
Introducció	n11
Factores	de riesgo de preeclampsia13
Algoritmo	o de la calculadora de <i>The Fetal Medicine Foundation</i> (FMF)15
Justificac	ión del informe17
Objetivo .	
Material y m	nétodo19
Tipo de e	studio19
Búsqueda	a bibliográfica19
Selección	de los artículos
Evaluació	n del riesgo de sesgo en los estudios
Extracció	n y síntesis de los resultados21
Análisis e	stadístico21
Análisis	s de sensibilidad23
Resultados	25
Resultado	o de la búsqueda25
Evaluació	on del riesgo de sesgo en los estudios27
Descripci	ón de los estudios29
Resulta	dos de rendimiento global de la prueba31
	Curvas ROC y áreas bajo la curva31
Tasas d	le detección
Análisis	s estadístico35
	Modelo 1: Validez diagnóstica del algoritmo FMF para el riesgo "a priori" de PE pretérmino35
	Modelo 2 Validez diagnóstica del Mini test de la FMF para el riesgo de PE pretérmino
	Modelo 3: Validez diagnóstica del Triple test de la FMF para el riesgo de PE pretérmino
	Análisis de sensibilidad
	Modelo 4: Validez diagnóstica del Cuádruple test de la FMF para el riesgo de PE pretérmino43

Análisis de sensibilidad	46
Discusión	49
Conclusiones	55
Referencias	57
Anexos	64
Anexo 1: estrategias de búsqueda	64
Anexo 2: estudios excluidos tras la lectura a texto completo	66
Anexo 3: Características de los estudios incluidos en la revisión siste	emática 67
Anexo 5: Análisis del efecto umbral en la estimación del riesgo "a pr Modelo 1	
Anexo 6: Análisis del efecto umbral en la estimación del riesgo Mod- Mini test	
Anexo 7: Análisis del efecto umbral en la estimación del riesgo de Modelo 3	•
Anexo 8: Análisis de sensibilidad del Triple test. Modelo 3	80
Por poblaciones	80
EUROPEA	80
NO EUROPEA	80
Con PAPP-A	80
Con PIGF	81
Por definición de PE	81
Definición PE 2001	81
Definición 2014	81
Anexo 9: Rendimiento del triple test para la PE temprana	82
Anexo 10: Análisis del efecto umbral en la estimación del riesgo de test. Modelo 4.	
Anexo 11: Análisis de sensibilidad del Modelo4: Cuádruple test	85
Por poblaciones	85
EUROPEA	85
NO EUROPEA	85
Por definición de PE	85
Definición PE 2001	85
Definición PE 2014	86
Anexo 12: Rendimiento del cuádruple test para la PE temprana	87

Resumen ejecutivo

Título: Validez diagnóstica del algoritmo de The Fetal Medicine Foundation para el cribado de preeclampsia pretérmino durante el primer trimestre de embarazo. Revisión sistemática y metaanálisis.

Autora: Isabel M. Martínez Férez

Documentación: Antonio Romero Tabares

ANTECEDENTES Y JUSTIFICACIÓN

La preeclampsia (PE) es una de las principales causas de morbilidad y mortalidad materna y fetal que afecta del 2 al 10 % de los embarazos. La administración profiláctica de aspirina reduce la incidencia de PE pretérmino (edad gestacional al momento del parto < 37 semanas) y las complicaciones asociadas, por lo que la identificación temprana de embarazos de alto riesgo es clave para reducir la carga de PE al permitir la orientación precisa de la aspirina profiláctica. La importancia del problema de salud y la existencia de un tratamiento preventivo, han llevado a recomendar el cribado de PE en el primer trimestre de gestación. Sin embargo, existe incertidumbre sobre la prueba más adecuada para el cribado. Una de las pruebas propuestas es el algoritmo de la FMF, desarrollado y validado principalmente en poblaciones procedentes del Reino Unido. Es fundamental establecer si el algoritmo presenta rendimientos diagnósticos similares en poblaciones diferentes a las que sirvieron para su diseño.

OBJETIVO

Valorar el rendimiento del algoritmo de la Fetal Medicine Foundation (FMF) como prueba de cribado de preeclampsia durante el primer trimestre de gestación en diferentes poblaciones de mujeres no seleccionadas con embarazo de un solo feto.

MATERIALES Y MÉTODOS

Se ha llevado a cabo una revisión sistemática de la literatura, con búsquedas bibliográficas en las bases de datos MedLine, Embase y Cochrane Library, hasta julio y agosto de 2021. Se ha realizado una lectura crítica de los artículos seleccionados, con el fin de identificar los problemas metodológicos que pudieran influir en su validez. Se ha llevado a cabo la combinación estadística de los datos mediante un metanálisis para la estimación de los valores ponderados de los parámetros de validez diagnóstica. Se ha abordado cuatro metanálisis diferentes, atendiendo a cuatro modelos resultantes de la combinación de diferentes factores de riesgo utilizados en el algoritmo de predicción de la FMF y se examinó la robustez de los hallazgos mediante un análisis de sensibilidad. Los modelos estudiados fueron: Modelo 1: riesgo "a priori" [características maternas (CM)]. Modelo 2: MINI TEST [CM, el factor biofísico MAP y un factor bioquímico (PIGF o PAPP-A)]. Modelo 3: CUÁDRUPLE TEST [CM, los factores biofísicos (MAP y UtAPI) y un factor bioquímico (PIGF o PAPP-A)].

RESULTADOS

La búsqueda bibliográfica ha permitido identificar 16 estudios prospectivos que valoraban el rendimiento diagnóstico del algoritmo de la FMF para el cribado de la PE

pretérmino. Los 16 estudios incluidos han analizado poblaciones procedentes de 18 países diferentes de América, Australia, Europa y Asia; con un rango de prevalencia de PE pretérmino entre 0,44% y 2,92%. El algoritmo de la FMF mostró un rendimiento diagnóstico global aceptable para el "riego a priori" con AUC entre 0,736 y 0,790; y bueno para el Triple y Cuádruple Test con AUC entre 0,787-0,948 y 0,843 y 0,96, respectivamente. El riesgo individual y la exactitud con la que el algoritmo de la FMF clasificaba un embarazo se determinó mediante la tasa de detección; esta tasa se estimó en todos los casos estableciendo el 10% de falsos positivos como punto de corte. Utilizando únicamente las características maternas las tasas de detección observadas estaban entre 39%-58% y eran similares para la PE temprana (<32-34 semanas) y para la PE pretérmino. El Triple Test y Cuádruple Test mostraron unas tasas de detección superiores a las anteriores alcanzando en algunos estudios el 80% para la PE pretérmino y hasta más de un 90% para la PE temprana. En todos los casos las tasas de detección estimadas presentaban unos intervalos de confianza del 95% muy amplios lo que indicaba una falta de precisión en estas estimaciones.

En el análisis estadístico realizado ningún modelo mostró efecto umbral como posible fuente de heterogeneidad. A una especificidad preestablecida del 90%, los cuatro modelos mostraron las siguientes sensibilidades ponderadas: "riesgo a priori" S=43% (IC 95% 39%-47%), Mini Test S=53 % (IC 95% 48%-59%), Triple Test S=70% (IC 95% 67%-74%) y Cuádruple Test S=74% (IC 95% 70%-77%). La comparación de las sensibilidades ponderadas mostró que las diferencias entre los modelos eran significativas excepto entre el triple Test y el Cuádruple Test. Los análisis de sensibilidad mostraron que ni la utilización de diferentes criterios de diagnóstico de PE ni la utilización del PIFG o PAPP-A afectaban a la sensibilidad ponderada mientras que la procedencia de la población por el contrario sí lo hacía. Así en el triple Test la sensibilidad ponderada para la población europea era de 77% (IC 96% 73%-81%) mientras que para otras poblaciones (asiática, australiana y americana) la sensibilidad ponderada era del 63% (IC 95% 58%-68%). En el Cuádruple Test la sensibilidad ponderada para la población europea era de 79% (IC 96% 75%-83%) mientras que para otras poblaciones (asiática y americana) la sensibilidad ponderada era del 63% (IC 95% 56%-70%).

Por último, el Triple Test de la FMF ha mostrado una mayor sensibilidad ponderada para PE temprana S=84% (IC 95% 78%-89%) que para la PE pretérmino S= 70% (IC 95% 67%-74%).

CONCLUSIONES

- 1. El algoritmo de la FMF para la identificación del riesgo alto de PE pretérmino presenta un buen rendimiento diagnóstico global.
- 2. La incorporación de los biomarcadores (biofísicos y bioquímicos) a las características maternas incrementa el rendimiento en el diagnóstico del riesgo de PE pretérmino.
- 3. El triple test de la FMF, que incluye las características maternas, MAP, UtAPI y PPAP-A o PIGF, es la prueba con mayor capacidad de discriminación. La incorporación de los dos biomarcadores PAPP-A y PIGF no incrementa la capacidad de discriminación de la prueba.
- 4. El Triple Test de la FMF ha mostrado su capacidad de discriminación en diferentes poblaciones, aunque su rendimiento se ve afectado por las características de las mismas. El rendimiento es mayor en población europea que en poblaciones procedentes de otras áreas

Abreviaturas

ACOG: American College of Obstetricians and Gynecologists.

AUC: área bajo la curva.

CM: características maternas.

CP+: coeficiente de probabilidad positivo.

CP-: coeficiente de probabilidad negativo.

E: especificidad.

FN: falsos negativos.

FP: falsos positivos.

ISSHP: International Society for the Study of Hypertension in Pregnancy .

MoM: múltiplo de la mediana.

MAP: presión arterial media.

NICE: National Institute for Health and Care Excellence.

PAPP-A: proteína plasmática A asociada al embarazo.

PE: preeclampsia.

PE < ns: PE que conduce al parto antes de la semana *n*.

 $PE \ge ns$: PE que conduce al parto en o después de la semana n.

PIGF o PLGF: el factor de crecimiento placentario.

S: sensibilidad.

sFLT-1: Tirosina quinasa 1 soluble tipo fms, también conocida como receptor 1 del factor de crecimiento vascular.

UtAPI o IPUT: índice de pulsatilidad de la arteria uterina.

VPN: valor predictivo negativo.

VPP: valor predictivo positivo.

Glosaario

- Curva de rendimiento diagnóstico (ROC, acrónimo del inglés *Receiver Operating Characteristic*): la curva ROC es un gráfico en el que se observan los pares de sensibilidad/especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados. En el eje Y de coordenadas se sitúa la sensibilidad o fracción de verdaderos positivos. En el eje X se sitúa la fracción de falsos positivos o 1-especificidad, calculada en el subgrupo no afectado.
- **Coeficiente de probabilidad positivo**: representa cuánto más probable es tener un positivo en un enfermo que en un sano.
- **Coeficiente de probabilidad negativo**: representa cuánto más probable es encontrar un test negativo en un enfermo que en un sano.
- **Efectividad**: beneficio, en términos de resultados en salud, que se deriva del uso de una tecnología o intervención para un determinado problema, en condiciones habituales (como el contexto de la atención sanitaria en los centros de salud u hospitales del sistema sanitario). También podría definirse como la evaluación o medida de cómo una intervención en salud alcanza sus objetivos en la práctica.
- **Eficacia**: beneficio derivado del uso de una tecnología o intervención para un determinado problema, en condiciones idóneas (como el contexto de un ensayo controlado aleatorio rigurosamente gestionado).
- **Especificidad** (de una prueba diagnóstica): probabilidad de que una prueba clasifique correctamente a una persona sana, es decir, la probabilidad de que para un sujeto sano se obtenga en una prueba diagnóstica un resultado negativo. La fórmula de la especificidad: Resultado Verdadero Negativo (VN)/ (Resultado Verdadero Negativo (VN)+ Resultado Falso Positivo (FP).
- **Falso negativo**: individuos clasificados por la prueba como no enfermos o de bajo riesgo pero que realmente estaban enfermos o eran de alto riesgo.
- **Falso positivo**: individuos clasificados por la prueba como enfermos o de alto riesgo pero que realmente no estaban enfermos o eran de bajo riesgo.
- **Metaanálisis**: es un estudio en el que se realiza la integración estructurada y sistemática de la información obtenida en diferentes ensayos clínicos, sobre un problema de salud determinado con el fin de dar una estimación cuantitativa sintética de todos los estudios disponibles.
- **Nulípara** (mujer): mujer que nunca ha parido, bien por no haber estado embarazada o por tener un embarazo que no ha pasado de 20 semanas.
- Sensibilidad (de una prueba diagnóstica): Es la capacidad de una prueba para detectar la enfermedad, es decir, la probabilidad de que para un sujeto enfermo se obtenga en una prueba diagnóstica un resultado positivo. La fórmula para calcularla es: Resultado Verdadero Positivo (VP) / Resultado Verdadero Positivo (VP)+ Resultado Falso Negativo (FN).
- PE temprana: preeclampsia que conduce al parto antes de las 32-34 semanas de gestación.

- PE pretérmino: preeclampsia que conduce al parto antes de las 37 semanas de gestación.
- **PE a término**: preeclampsia que conduce al parto a partir de las 37 semanas de gestación.
- **Tasa de detección**: proporción de casos identificados en el cribado como positivos que finalmente desarrollan la enfermedad [Nº casos / nº total de casos estimados]. Es una medida de sensibilidad del cribado.
- Validez (de una prueba diagnóstica): Grado de concordancia entre el resultado de la prueba y el diagnóstico clínico definitivo. Normalmente se mide en función de la sensibilidad y la especificidad.
- **Valor predictivo negativo**: probabilidad de que un individuo con un resultado negativo en la prueba no tenga la enfermedad en estudio.
- **Valor predictivo positivo**: probabilidad de que un individuo con un resultado positivo en la prueba tenga la enfermedad en estudio.
- **Verdadero negativo**: individuos clasificados por la prueba como no enfermos o de bajo riesgo y que realmente no lo eran.
- **Verdadero positivo**: individuos clasificados por la prueba como enfermos o de alto riesgo y que realmente lo eran.

Introducción

La preeclampsia (PE) es un trastorno hipertensivo de carácter progresivo y relativamente común que ocurre durante el embarazo, tiene múltiples subtipos y sus resultados de salud son potencialmente graves, incluso fatales. La etiología específica de la preeclampsia no está clara (USPSTF, 2017). La PE ha sido definida como hipertensión de nueva aparición (o, en pacientes con hipertensión existente, empeoramiento de la hipertensión) que ocurre después de las 20 semanas de gestación, combinada con proteinuria de nueva aparición (exceso de proteína en la orina) u otros signos o síntomas que involucran múltiples sistemas orgánicos. En 2018 la *International Society for the Study of Hypertension in Pregnancy* (ISSHP) actualizó su definición quedando como se muestra a continuación (Brown, Magee, Kenny, Karumanchi, McCarthy, Saito *et al.*, 2018).

La preeclampsia es hipertensión gestacional acompañada de la aparición *de novo* a partir de las 20 semanas de gestación de una o más de las siguientes condiciones:

- 1. Proteinuria
- 2. Otra disfunción orgánica materna, que incluye:
 - Insuficiencia renal aguda (AKI) (creatinina ≥90 μmol/L; 1 mg/dL)
 - compromiso hepático (transaminasas elevadas, p. ej., ALT o AST > 40 UI/L) con o sin dolor abdominal epigástrico o en el cuadrante superior derecho
 - complicaciones neurológicas (los ejemplos incluyen eclampsia, estado mental alterado, ceguera, accidente cerebrovascular, clonus, cefaleas de intensidad severa, escotomas visuales persistentes)
 - Complicaciones hematológicas (trombocitopenia recuento de plaquetas por debajo de 150.000/μL, coagulación intravascular diseminada (DIC), hemólisis)
- 3. Disfunción uteroplacentaria (como restricción del crecimiento fetal, análisis anormal de la forma de onda Doppler de la arteria umbilical, o el nacimiento de un feto muerto)

La preeclampsia puede empeorar rápidamente y sin previo aviso. Entre los efectos adversos perinatales para el feto y el recién nacido se incluyen la restricción del crecimiento intrauterino, un bajo peso al nacer y la muerte fetal. Las complicaciones asociadas con la preeclampsia conducen a la inducción temprana del trabajo de parto o al parto por cesárea y posterior parto prematuro (USPSTF, 2017).

Aproximadamente el 90 por ciento de los casos se presentan en el período prematuro tardío (≥34 a <37 semanas), en el periodo de término (≥37 a <42 semanas) o en el posparto (≥42 semanas) y en estos casos se tienen buenos resultados maternos, fetales y neonatales, aunque no hay que olvidar que tanto la madre como el niño se encuentran en una situación de alto riesgo de morbilidad o mortalidad. El 10 por ciento restante de los casos tiene una presentación temprana (<34 semanas) y lleva asociado un alto riesgo de parto prematuro, muy prematuro o extremadamente prematuro. A largo plazo, las pacientes con preeclampsia presentan un mayor riesgo de desarrollar enfermedades cardiovasculares y renales (August y

Sibai, 2021). La aparición temprana de preeclampsia se asocia con varias enfermedades placentarias siendo la más común la disfunción vascular placentaria caracterizada por un tamaño placentario reducido, infarto multifocal, hemorragia y alteraciones en las arterias espirales del útero (Bujold, Chaillet y Kingdom, 2019).

A nivel mundial, la incidencia de preeclampsia oscila entre 2-10% de los embarazos, la cual es precursor de la eclampsia y varía en todo el mundo. La Organización Mundial de la Salud (OMS) estima que la incidencia de preeclampsia es siete veces mayor en los países en desarrollo que en los desarrollados (2,8% y 0,4% de los nacidos vivos, respectivamente). La incidencia de eclampsia en los países desarrollados de Norteamérica y Europa es similar y se estima alrededor de 5 a 7 casos por cada 10.000 partos, mientras que en países en desarrollo es variable, oscilando entre 1 caso por cada 100 embarazos a 1 por cada 1.700 embarazos (Vargas, Acosta y Moreno, 2012). Cada año la preeclampsia es responsable de más de 500.000 muertes fetales y neonatales y de más de 70.000 muertes maternas (Brown *et al.*, 2018); es la segunda causa principal de mortalidad materna y puede provocar otras complicaciones maternas graves, como accidentes cerebrovasculares, eclampsia e insuficiencia orgánica.

Prevención de la preeclampsia

El diagnóstico temprano seguido de un tratamiento adecuado, incluido el parto, puede prevenir algunas de las secuelas graves de la enfermedad, como las convulsiones eclámpticas y la insuficiencia multiorgánica. Dado que no existe un tratamiento curativo que no sea el parto, una intervención que pudiera prevenir la preeclampsia tendría un impacto significativo en la salud materna e infantil. Se han investigado, en ensayos aleatorizados, muchas estrategias diferentes para prevenir la preeclampsia y no es sorprendente que la mayoría de los enfoques simples no hayan tenido éxito dada la complejidad de la patogénesis y la probabilidad de que múltiples etiologías o vías den como resultado el síndrome clínico.

La profilaxis con aspirina en dosis bajas es la intervención farmacológica preventiva más extendida. La aspirina en bajas dosis, cuando se administra a mujeres con riesgo de la enfermedad, reduce la frecuencia de la preeclampsia y los resultados adversos relacionados con el embarazo (parto prematuro, restricción del crecimiento y muerte fetal o neonatal). En general, la administración de agentes antiplaquetarios a 1000 mujeres dio lugar a 20 embarazos menos con resultados adversos graves. Además, la aspirina en baja dosis parece ser un tratamiento seguro sin daños evidentes (Duley, Meher, Hunter, Seidler y Askie, 2019; Henderson, Vesco, Senger, Thomas y Redmond, 2021). La identificación de las mujeres que tienen más probabilidades de responder a la aspirina en dosis bajas mejoraría la orientación del tratamiento (Duley *et al.*, 2019).

Este perfil de seguridad materno-fetal durante el embarazo; ha llevado a considerar el tratamiento con aspirina en bajas dosis una estrategia preventiva razonable para las mujeres con riesgo de preeclampsia. El tratamiento temprano (antes de las 16 semanas) parece ser importante, ya que las características fisiopatológicas de la preeclampsia se desarrollan al principio del embarazo, semanas antes de que la enfermedad clínica sea aparente (August y Jeyabalan, 2021). Sin embargo, tanto la dosis como el momento más adecuado para el inicio del tratamiento preventivo con aspirina han sido aspectos controvertidos. Con respecto a la dosis, se había observado que la prevención de la preeclampsia y la restricción del crecimiento fetal con aspirina al comienzo del embarazo presentaba un efecto de dosis-respuesta, por lo que las dosis más altas de aspirina se asociaban a una mayor reducción de preeclampsia y restricción de crecimiento fetal (Roberge, Nicolaides, Demers, Hyett., Chaillet y Bujold, 2017) y por otro lado, en cuanto al momento adecuado de tratamiento, revisiones

sistemáticas sugerían el beneficio de la aspirina a baja dosis cuando se iniciaba el tratamiento a las 16 semanas de gestación o antes, o a partir de las 12 semanas. No estando claro si el inicio del tratamiento antes de las 12 semanas tendría beneficios adicionales sin un aumento de los efectos adversos (Roberge, Villa, Nocolaides, Giguère, Vaino, Bakthi *et al.*, 2012; Duley *et al.*, 2019).

El estudio ASPRE ha demostrado que la administración de aspirina en una dosis de 150 mg por día desde las semanas 11 a 14 de gestación hasta 36 semanas de gestación en embarazos con alto riesgo de PE reduce la tasa de PE pretérmino en un 62 % y no reduce la incidencia de PE a término. El estudio concluye que estos resultados respaldaban el inicio más temprano de la terapia con aspirina (Rolnik, Wright, Poon, O' Gorman, Syngelaki, Paco de Matalla *et al.*, 2017).

El uso de la aspirina en bajas dosis para la prevención de la preeclampsia en mujeres con alto riesgo está ampliamente contemplado y recomendado (Organización Mundial de la Salud, 2014; NICE, 2019; *US Preventive Services Task Force*, 2021 etc.).

Factores de riesgo de preeclampsia

El enfoque tradicional para identificar el riesgo de PE se basa en las características demográficas de la madre y de su historial médico. Una revisión sistemática con metanálisis de 92 estudios y que incluía 25.356.688 embarazos que evaluaba la asociación entre los factores de riesgo clínicos identificados antes de las 16 semanas de gestación y el riesgo de PE mostró entre los factores más relevantes los siguientes (Bartsch, Medcalf, Park, Ray, High Risk of Preeclampsia Identification Group, 2016):

- Los antecedentes de PE: RR 8,4 (IC 95% 7,1-9,9).
- Hipertensión crónica: RR 5,1 (IC 95% 4,0-6,5).
- Diabetes pregestacional: RR 3,7 (IC 95% 3,1-4,3).
- Embarazos multifetales: RR 2,9 (IC 95% 2,6-3,1).
- Síndrome de anticuerpos antifosfolípidos: RR 2,8 (IC 95% 1,8-4,3).
- Índice de masa corporal (IMC > 30): RR 2,8 (IC95% 2,6-3,1).
- Uso de tecnologías de reproducción asistida: RR 1,8 (IC 95% 1,6-2,1).
- Nuliparidad: RR 2,1 (IC 95% 1,9-2,4).
- Enfermedad renal crónica: RR 1,8 (IC95% 1,5-2,1).
- Lupus eritematoso: RR 2,5 (IC95% 1,0-6,3).

Este enfoque tradicional ha sido el adoptado por organismos como el *National Institute for Health and Care Excellence* (NICE) y la *American College of Obstetricians and Gynecologists* (ACOG) que han propuesto la detección de preeclampsia mediante un listado de factores de riesgo maternos en el que cada factor de riesgo es considerado esencialmente como una prueba de detección separada, estableciendo una tasa de detección aditiva y una tasa de detección positiva. La evidencia ha demostrado que el cribado de preeclampsia basado en el enfoque NICE y ACOG tiene un rendimiento subóptimo, ya que las recomendaciones del NICE solo logra tasas de detección del 41% y 34%, con una tasa de falsos positivos del 10%, para preeclampsia prematura y a término, respectivamente. El cribado

basado en las recomendaciones del 2013 de la ACOG muestra tasas de detección del 5% y 2% para PE pretérmino y PE a término, respectivamente, con una tasa de falsos positivos del 0,2%. Este enfoque presenta un rendimiento predictivo limitado que actualmente no es considerado suficiente para predecir la preeclampsia de manera eficaz (Chaemsaithong, Sahota y Poon, 2020).

Esta falta de rendimiento predictivo llevó al desarrollo de varios modelos de predicción basados en algoritmos obtenidos mediante análisis de regresión logística. Estos algoritmos permitían establecer el riesgo individual de PE de cada paciente mediante la combinación de los factores de riesgo maternos en lugar de considerarlos de forma separada. Un paso más allá en el desarrollo de modelos predictivos fue la aplicación del teorema de Bayes. Este nuevo enfoque combinaba el riesgo a priori, determinado por las características maternas y los antecedentes médicos y obstétricos, con los resultados de adecuados marcadores biofísicos y bioquímicos que permitieran una evaluación de los riesgos de preeclampsia específicos de cada paciente con el objetivo de establecer el momento más idóneo del parto antes de que aparezca la PE. En estos análisis los valores de los biomarcadores se transformaban en múltiplos de la mediana (MoM) para tener en cuenta los efectos de la edad gestacional y las características maternas, como el peso y el origen racial, asociados con los biomarcadores individuales. Además, el teorema de Bayes proporciona un marco natural para la predicción dinámica, en el que la evaluación de riesgos se actualiza a medida que se dispone de nueva información durante el embarazo. La incorporación de información de biomarcadores utilizando probabilidades significa que se pueden agregar nuevos marcadores ampliando un modelo existente en lugar de desarrollar un modelo completamente nuevo. Sobre la base del teorema de Bayes, se ha desarrollado un modelo de riesgo competitivo, que consiste en un modelo de tiempo de supervivencia para la edad gestacional en el que se establece el momento en el que tendría lugar un parto con preeclampsia (Chaemsaithong, Sahota y Poon, 2020).

El reto es por lo tanto identificar los predictores adecuados para diseñar el algoritmo. Los predictores que han mostrado mejor rendimiento de la prueba han sido el índice de masa corporal (IMC) >35 kg.m-2, con una especificidad del 92% (IC 95%, 89-95%) y una sensibilidad del 21% (IC 95%, 12-31%); IMC >25 kg.m-2, con una especificidad del 73% (IC 95%: 64-83%) y una sensibilidad del 47% (IC 95%: 33-61%); índice de pulsatilidad de la arteria uterina (UtAPI) en el primer trimestre o índice de resistencia > 90° percentil [especificidad del 93% (IC 95%: 90-96%) y sensibilidad del 26% (IC 95%: 23-31%)]; factor de crecimiento placentario (PIGF) [especificidad 89% (IC 95%, 89-89%) y sensibilidad 65% (IC 95%, 63-67%)]; y proteína placentaria 13 [especificidad 88% (IC 95%, 87-89%) y sensibilidad 37% (IC 95%, 33-41%)]. Ningún marcador por sí solo ha mostrado un rendimiento de la prueba adecuado para el uso clínico rutinario. Los modelos que combinan marcadores son prometedores, pero ninguno fue sometido a una validación externa (Townsend, Khalil, Premakumar, Allotey, Snell, Chan *et al.* 2019).

En resumen, los algoritmos incluyen de forma general los siguientes factores.

- Historial y características maternas
- 2. Marcadores bioquímicos: el factor de crecimiento placentario (PIGF), el receptor de tirosina quinasa 1 similar a fms soluble (sFLT-1), la proteína plasmática A asociada al embarazo (PAPP-A)
- 3. Marcadores biofísicos: la presión arterial media (MAP), el índice de pulsatilidad de la arteria uterina (UtAPI)

La mayoría de los algoritmos propuestos no han sido sometidos a validación externa; sin embargo, uno de los algoritmos que sí se ha sometido a una validación interna y externa ha sido el modelo de predicción del primer trimestre de la *Fetal Medicine Foundation* (FMF) (la denominada prueba triple), que combina factores de riesgo maternos y mediciones de la presión arterial media (MAP), el índice de pulsatilidad de la arteria uterina (UtAPI) y el factor de crecimiento sérico-placentario (PIGF) (Chaemsaithong, Sahota y Poon, 2020).

Algoritmo de la calculadora de *The Fetal Medicine Foundation* (FMF)

El modelo de predicción del primer trimestre de la *Fetal Medicine Foundation* (FMF) está disponible *online* en forma de calculadora del riesgo de preeclampsia y es una aplicación que utiliza el teorema de Bayes para combinar factores de riesgo previo y el historial médico materno con los resultados de varias mediciones biofísicas y bioquímicas con la finalidad de estimar el riesgo posterior de preeclampsia (https://fetalmedicine.org/research/assess/preeclampsia/background).

Como indica la *Fetal Medicine Foundation* el software de la calculadora de riesgo se basa en una investigación realizada por *The Fetal Medicine Foundation* e indica que ni la FMF ni ninguna otra parte involucrada en el desarrollo de este software serán responsables de los resultados producidos cuando se utilicen datos de fuentes no confirmadas. Asimismo, puntualizan que la evaluación del riesgo clínico requiere que las mediciones bioquímicas y de ultrasonido sean tomadas y analizadas por médicos y laboratorios acreditados.

La calculadora estima el riesgo de preeclampsia en tres periodos del embarazo y para cada uno de ellos considera factores de riesgo específicos.

- 11⁺⁰ a 14⁺¹ semanas: Los marcadores útiles considerados en este periodo son: la presión arterial media (MAP), el índice de pulsatilidad medio de la arteria uterina (UtAPI), factor de crecimiento placentario (PIGF) (o PAPP-A cuando no se dispone de PIGF).
- >19⁺⁰ a 24⁺⁶ semanas: MAP, UtAPI, PIGF y sFLT-1 (factor antiangiogénico *Serum soluble fms-like tyrosine kinase-1*)
- >30⁺⁰ a 37⁺⁶ semanas: MA, PLGF y sFLT-1, con un discreto beneficio al incluir UtAPI.

La calculadora propone como razonable para el cribado de PE en la semana 11^{+0} a 14^{+1} utilizar un umbral de riesgo de 1 en 150 para definir el grupo de alto riesgo que se beneficiaría del uso profiláctico de aspirina cuando se realice en un entorno con población predominantemente blanca. En una población blanca, la calculadora estima que para un corte de riesgo de 1 de cada 100 y 1 de cada 150, las tasas de cribado positivo son de aproximadamente el 10% y el 16% y las tasas de detección para la EP temprana (< 32 semanas) son del 88% y el 94% y para la EP prematura (<37 semanas) son del 69% y 81%, respectivamente.

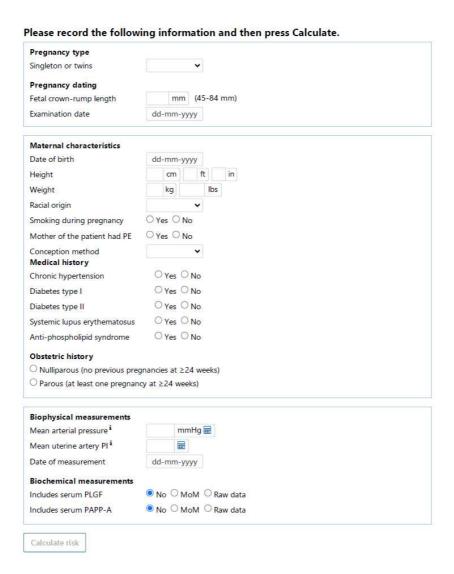
Sin embargo, la FMF aconseja que en mujeres de etnia negra este punto de corte de 1 en 150 debería ser menor, ya que para este este umbral la tasa de cribado positivo sería de alrededor del 40% con unas tasas de detección para la EP precoz y prematura del 100% y 95%, respectivamente. Esta diferencia de umbrales propuesta se debería al hecho de que la

prevalencia de la EP prematura es más de tres veces mayor en la etnia negra que en la blanca.

La calculadora de la FMF permite estimar el riesgo individual de sufrir preeclampsia a partir únicamente de las características maternas, lo que denomina "Riesgo *a priori*" y, a partir de la combinación de las características maternas y los factores biofísicos y bioquímicos, lo que se denomina "Triple test".

El ensayo ASPRE mostró que el uso del modelo de predicción de la FMF seguido de la administración de aspirina en dosis bajas reducía la tasa de preeclampsia pretérmino en un 62%. El número necesario para prevenir 1 caso de preeclampsia prematura mediante la prueba triple FMF era 250. La clave para mantener un rendimiento de detección óptimo es establecer protocolos estandarizados para las mediciones de los biomarcadores y la evaluación regular de la calidad de los mismos, ya que una medición inexacta puede afectar el rendimiento de la detección (Chaemsaithong, Sahota y Poon, 2020).

Figura 1. Calculadora de riesgo de preeclampsia de la FMF para las semanas 11⁺⁰-14⁺¹



Justificación del informe

Actualmente el riesgo de preeclampsia en el embarazo se valora en función de las características y el historial médico materno, pero esta valoración ha demostrado ser insuficiente a la hora de detectar de manera precoz la aparición de la enfermedad. Esta situación ha llevado al desarrollo de numerosos algoritmos, ente ellos el propuesto por la FMF, que intentan establecer con mayor precisión el riesgo de preeclampsia.

El algoritmo de la FMF ha sido desarrollado y validado principalmente en poblaciones procedentes del Reino Unido. Se ha observado que los parámetros biofísicos y bioquímicos utilizados en el mismo para establecer el riesgo de preeclampsia, e incluso los umbrales para su detección, podrían verse afectados por algunas características como por ejemplo la etnia. Por consiguiente, es fundamental establecer si el algoritmo presenta rendimientos diagnósticos similares en otras poblaciones, es decir, si en otras poblaciones diferentes a las que sirvieron para su diseño su aplicación dará resultados similares.

Objetivo

El objetivo del informe ha sido valorar el rendimiento del algoritmo de FMF como prueba de cribado de preeclampsia durante el primer trimestre de gestación en mujeres no seleccionadas con embarazo de un solo feto.

Este informe trata de dar respuesta a la siguiente pregunta de investigación:

¿El cribado mediante el algoritmo de la FMF durante el primer trimestre de embarazo para la detección del riesgo de preeclampsia pretérmino en mujeres con embarazo de un solo feto permite identificar los embarazos con alto riesgo que podrían beneficiarse del uso profiláctico de aspirina?

Material y método

Tipo de estudio

Revisión sistemática de la literatura siguiendo las recomendaciones recogidas en la declaración PRISMA para la publicación de revisiones sistemáticas y metaanálisis (Urrutia y Bonfill, 2010).

Búsqueda bibliográfica

Se ha realizado una búsqueda bibliográfica en las bases de referencias MEDLINE (mediante OVID) y EMBASE. También, se consultó la *Cochrane Library* para la identificación de revisiones sistemáticas. Estas búsquedas se realizaron hasta julio y agosto de 2021 y las estrategias diseñadas para cada una de las bases se han descrito en el Anexo 1.

Además, la búsqueda bibliográfica se completó con una revisión secundaria de la bibliografía de los artículos obtenidos en las estrategias antes descritas.

Selección de los artículos

La selección de los artículos relevantes que podían dar respuesta a los objetivos de este informe se realizó siguiendo los criterios de inclusión y exclusión establecidos *a priori* y descritos a continuación.

Los criterios de inclusión.

- Población: mujeres embarazadas nulíparas o multíparas con embarazo de un solo feto y sin riesgo previo de preeclampsia.
- Intervención: algoritmo de la FMF para el cribado de preeclampsia durante el primer trimestre de embarazo.
- Comparación: Seguimiento y resultado del embarazo.
- Resultados de rendimiento de la prueba medido como número de casos de preeclampsia pretérmino (<34 semanas o <37 semanas) y los parámetros
 estadísticos de validez diagnóstica (tasa de detección, sensibilidad,
 especificidad, valores predictivos positivo y negativo, coeficientes de
 probabilidad (CP) y AUC).

Los criterios de exclusión.

- Embarazos múltiples.
- Embarazos seleccionados con alto riesgo de preeclampsia.
- Idioma diferente a inglés, español, italiano, portugués y francés.
- Estudios con diseño retrospectivo, de caso-control.
- Publicaciones como comunicaciones a congresos, cartas, editoriales,

- comentarios, revisiones no sistemáticas y todas aquellas que no incluyeran datos originales sobre el tema.
- Estudios en los que no fuera posible extraer la información necesaria para realizar la tabla 2x2 de clasificación cruzada.

La selección de los estudios ha sido llevada a cabo por un único investigador.

Evaluación del riesgo de sesgo en los estudios

Se ha llevado a cabo una lectura crítica de los artículos seleccionados con el objeto de identificar los problemas metodológicos que pudieran afectar a la validez interna y externa de los estudios.

La valoración del riesgo de sesgos, siguiendo los criterios de la herramienta Quadas-2 para pruebas diagnósticas, y la representación del control de sesgos se han realizado mediante el programa *Review Manager (RevMan) [Computer program]. Version 5.4.1, The Cochrane Collaboration, 2020.* Siguiendo los criterios de la herramienta Quadas-2 se valoró el riesgo de sesgos en los cuatro dominios recogidos en ella: selección de pacientes, prueba diagnóstica de estudio, prueba diagnóstica de referencia y el flujo y tiempo de realización de las dos pruebas.

- Dominio selección de pacientes: para la valoración de sesgo en la selección de pacientes sólo se consideró un riego alto cuando no fuera posible descartar el diseño de caso-control. Se consideró incierto el riesgo de sesgo si el estudio no explicitaba los criterios de inclusión y exclusión de las pacientes o las pérdidas de pacientes elegibles.
- Dominio prueba de estudio: se consideró al valorar este dominio que tratándose de una prueba matemática cuyo resultado es un número no presentaría en principio riesgo de sesgo alto siempre que se explicitara el umbral considerado para la clasificación del riesgo de las pacientes. Sin embargo, se consideró el riego incierto si dicho umbral no era explícito.
- Dominio prueba de referencia: la prueba de referencia al ser el resultado del embarazo y la identificación de los casos de PE se consideró que el sesgo sería bajo siempre que se explicitara claramente los criterios establecidos para diagnosticar la enfermedad., en caso de que no se explicitaran claramente o no fueran explicitados en absoluto, se consideraría el riesgo de sesgo incierto o alto, respectivamente.
- Dominio flujo y tiempo: en este caso se valoró las pérdidas de pacientes y que las mediciones necesarias para la prueba de estudio se realizaran en el primer trimestre, ya que el tiempo entre las dos pruebas era necesariamente la duración del embarazo.

La lectura crítica de los estudios y la valoración del riesgo de sesgos han sido llevadas a cabo por un único investigador.

Extracción y síntesis de los resultados

La extracción de los principales resultados descritos en cada estudio se ha resumido en tablas de evidencia siguiendo un modelo desarrollado y diseñado previamente. Las variables recogidas incluyeron información general como el autor, el país, el año de publicación, los objetivos, las características de los pacientes, así como de la intervención y el seguimiento.

Los resultados de interés para la valoración del rendimiento de la prueba fueron:

- Las áreas bajo la curva (AUC), parámetro que proporciona una medida del funcionamiento global de una prueba diagnóstica. En este caso proporciona una medida de la capacidad global del algoritmo de la FMF de discriminar entre riesgo alto o bajo de PE pretérmino, pero no permite estimar el riesgo individual. Los valores que puede tomar el parámetro van desde 0,5 (prueba sin capacidad de discriminar) hasta 1 (prueba que discrimina perfectamente entre alto y bajo riesgo). A mayor valor del AUC mayor capacidad de discriminación de la prueba.
- Parámetros de exactitud diagnóstica: Tasa de detección, sensibilidad, especificidad, VPP, VPN, CP+, CP-. Estos parámetros permiten estimar el riesgo individual y la exactitud con la que se clasifica el riesgo de un paciente determinado. En aquellos estudios dónde no se aportaban estos parámetros diagnósticos pero fue posible construir la tabla 2x2 de clasificación cruzada dichos parámetros fueron estimados por la autora de esta revisión utilizando la calculadora de Fisterra (https://www.fisterra.com/mbe/investiga/pruebas_diagnosticas.asp). En aquellos casos donde no fue posible extraer los datos de la tabla 2x2, el estudio no se incluyó en el análisis.

La extracción y la síntesis de los resultados han sido realizados por un único investigador.

Análisis estadístico

Se valoró la realización de la combinación estadística de los datos mediante un metanálisis al considerar que los estudios incluidos en la revisión presentaban el mismo diseño, habían evaluado la misma prueba diagnóstica, usando un mismo método de referencia y en pacientes similares (Zamora Romero, Nieves Plana y Abraira Santos, 2009). El objetivo de este metanálisis era calcular un estimador ponderado de los efectos. Se ha considerado necesario un mínimo de 3 estudios para la combinación estadística.

Se han abordado cuatro metaanálisis diferentes, atendiendo a los factores utilizados en el modelo de predicción de la FMF:

- Modelo 1: calcula el riesgo de PE con las características maternas (CM), riesgo a priori.
- Modelo 2: calcula el riesgo de PE con las CM, el factor biofísico MAP y un factor bioquímico (PIGF o PAPP-A), se corresponde al denominado MINI TEST. Este modelo de predicción se ha elegido ya que su introducción no implicaría un coste

adicional al incluir pruebas que se realizan ya en nuestro sistema sanitario dentro del control normal del embarazo.

- Modelo 3: calcula el riesgo de PE con las CM, los factores biofísicos (MAP y UtAPI) y un factor bioquímico (PIGF o PAPP-A), se corresponde al denominado *Triple Test*.
- Modelo 4: calcula el riesgo de PE con las CM, los factores biofísicos (MAP y UtAPI) y los factores bioquímicos (PIGF y PAPP-A), se corresponde al denominado Cuádruple Test.

En primer lugar, se exploró la existencia o ausencia de efecto umbral como posible fuente de heterogeneidad que podría suponer la utilización de diferentes umbrales para clasificar el resultado de la prueba como positivo (riesgo alto) o negativo. Esta exploración se llevó a cabo mediante dos técnicas

- Una representación gráfica de las parejas de sensibilidad y especificidad de cada estudio en el plano ROC que seguirían un patrón curvilíneo en caso de efecto umbral, y
- el **coeficiente de Spearman** entre sensibilidad y especificidad que en caso de existir dicho efecto mostraría una correlación inversa.

En el caso de no haber efecto umbral, se analizó la presencia de heterogeneidad entre los estudios mediante la prueba *chi-cuadrado* para la sensibilidad y especificidad y la prueba Cochran-Q para los coeficientes de probabilidad (CP). Para cuantificar la magnitud de la heterogeneidad se usó el índice de inconsistencia I^2 (valores inferiores al 25% indican baja heterogeneidad, alrededor del 50- 75% moderada y al 75% alta) (Zamora Romero, Plana y Abraira Santos, 2009). La significación estadística se estableció en p < 0.05. Los criterios para la valoración de los CP fueron (Jaeschke, Guyatt y Sackett,1994):

CP+	CP-	Valor de la prueba
>10	< 0,1	Prueba con muy buena capacidad diagnóstica.
5-10	0,1-0,2	Prueba con capacidad diagnóstica moderada
2-5	0,2-0,5	Prueba de aportación dudosa
1-2	0,5-1	Prueba sin utilidad diagnóstica

Se abordó la agregación cuantitativa de datos (metaanálisis) con el fin de obtener una estimación promedio y su IC al 95% de los parámetros estadísticos sensibilidad, especificidad, CP+, CP-. Se utilizó un modelo de efectos fijos en ausencia de heterogeneidad entre los estudios, este modelo asume que las variaciones son debidas al azar; por el contrario, se utilizó un modelo de efectos aleatorios si existía heterogeneidad entre los estudios elegibles, este modelo suele dar un IC más amplio que el modelo de efectos fijos. Los

métodos estadísticos, fueron para el modelo de efectos fijos el de Mantel-Haenszel y para el modelo de efectos aleatorios el de DerSimonian-Laird. Los índices promedios se calcularon ponderados por la inversa de la varianza.

Análisis de sensibilidad

En los casos en los que el estadístico mostrara heterogeneidad, se examinó la robustez de los hallazgos mediante un análisis de sensibilidad. En concreto, se analizaron por subgrupos los resultados en función a varias variables que podrían explicar esa variación siempre que hubiesen más de tres estudios para la estimación global de los parámetros diagnóstico. Las variables utilizadas en los análisis de sensibilidad han sido:

- Procedencia de la población.
- Los criterios de diagnóstico de PE utilizados: criterios antiguos y actualizados
- Los parámetros del algoritmo usados (triple test con PIGF y triple test con PAPP-A).

El análisis estadístico de heterogeneidad, combinación de resultados y de subgrupos (análisis de sensibilidad) se llevó a cabo el programa Meta-Disc versión 1.3. Se valoró las diferencias entre los parámetros diagnósticos mediante la comparación de proporciones para muestras independientes (programa EPIDAT). La significación estadística se estableció en p < 0.05.

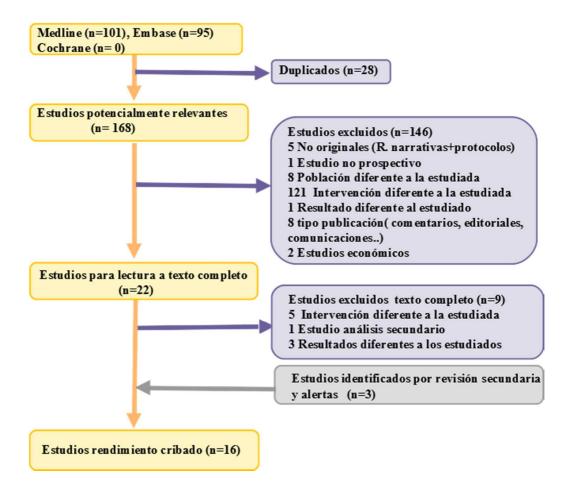
Resultados

Resultado de la búsqueda

Las estrategias de búsqueda identificaron un total de 196 artículos, 28 de ellos 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 146 que no cumplían los criterios de inclusión establecidos y se seleccionaron 22 referencias para su lectura a texto completo. De ellas se incluyeron 13 en el informe y las otras nueve fueron excluidas, estas referencias junto a sus motivos de exclusión se detallan en el Anexo 2.

A las 13 referencias identificadas en las estrategias se sumaron tres identificadas por revisión secundaria y sistema de alertas. La Figura 1 muestra el proceso de selección de los artículos incluidos en nuestra revisión sistemática y metaanálisis.

Figura 2. Diagrama de flujo de selección de referencias



En total se seleccionaron 16 estudios que valoraban el rendimiento diagnóstico del algoritmo de la FMF para el cribado de la PE, estos estudios ordenados por fecha de publicación se han recogido en la Tabla 1.

Tabla 1. Estudios incl i	uidos en la revis	ión para la valoración del rendir	miento del algoritmo de la FMF
Estudio	País	Tipo de estudio	Objetivo estudio
Park <i>et al.</i> , 2013	Australia	Cohorte prospectiva multicéntrica (2 hospitales de Sydney)	Validación externa del algoritmo FMF
Oliveira, Magder, Blitzer y Baschat, 2014	E.E.U.U.	Cohorte prospectiva multicéntrica (4 hospitales de Baltimore)	Validación externa del algoritmo FMF
Skrästad, Hov, Blaas,	Noruega	Cohorte prospectiva	Validación externa del algoritmo FMF en
Romundstad y Salvesen, 2015		unicéntrica (1 hospital de Trondheim)	una población nulípara
Mosimann <i>et al.</i> , 2017	Suiza	Cohorte prospectiva unicéntrica (1 hospital de Berna)	Validación externa del algoritmo FMF
Rolnik <i>et al.,</i> 2017	Internacional Europea	Cohorte prospectiva multicéntrica (13 hospitales europeos)	Validación externa del algoritmo FMF
O´Gorman <i>et al</i> .,	Internacional	Cohorte prospectiva	Validación externa del algoritmo FMF y
2017a	Europea	multicéntrica (12 hospitales europeos)	comparación con NICE y ACOG
Guizani <i>et al</i> ., 2018	Bélgica	Cohorte prospectiva unicéntrica (1 hospital de Bruselas)	Validación externa del algoritmo FMF
Tan <i>et al.,</i> 2018	Reino Unido	Cohorte prospectiva multicéntrica (Varios hospitales de Inglaterra)	Validación externa del algoritmo FMF y comparación con NICE
Di Martino <i>et al</i> ., 2019	Italia	Cohorte prospectiva multicéntrica (4 hospitales del norte de Italia)	Validación externa del algoritmo FMF
Lobo <i>et al.</i> , 2019	Brasil	Cohorte prospectiva	Validación externa del algoritmo FMF
Chaemsaithong <i>et</i> <i>al.</i> ,2019	Internacional Asiática	Cohorte prospectiva multicéntrica (11 hospitales asiáticos)	Validación externa del algoritmo FMF
Boutin, Guerby, Gasse, Tapp y Bujold, 2021	Canadá	Cohorte prospectiva	Validación externa del algoritmo FMF en una población nulípara
Goto <i>et al.</i> , 2021	Japón	Cohorte prospectiva unicéntrica (1 hospital de Tokio)	Validación externa del algoritmo FMF
Chen <i>et al.</i> , 2021	Taiwán	Cohorte prospectiva unicéntrica (1 hospital de Taiwan)	Validación externa del algoritmo FMF
Hu <i>et al.</i> , 2021	China	Cohorte prospectiva multicéntrica (13 hospitales chinos)	Validación externa del algoritmo FMF

Tabla 1. Estudios incluidos en la revisión para la valoración del rendimiento del algoritmo de la FMF				
Estudio	País	Tipo de estudio	Objetivo estudio	
Prasad <i>et al.</i> , 2021	India	Cohorte prospectiva unicéntrica (1 hospital de Nueva Delhi)	Validación externa del algoritmo FMF	

Evaluación del riesgo de sesgo en los estudios

La mayoría de los estudios seleccionados utilizaron los datos procedentes de cohortes prospectivas, pero el análisis de dichos datos fue realizado de manera retrospectiva una vez terminados los seguimientos de los embarazos y se disponía del resultado de la prueba de referencia (resultado del embarazo y aparición o no de preeclampsia). En general, los estudios incluidos presentaban una buena validez interna ya que su diseño y realización estaban bien llevados a cabo y los resultados eran extraídos correctamente de los participantes del mismo.

Sesgo en el dominio de población

Los estudios realizaron un cribado poblacional durante el primer trimestre de gestación a las mujeres con embarazo único que llegaron a los centros sanitarios para su revisión rutinaria. Aunque en la mayoría de los casos el cribado se realizó específicamente para la PE, en un estudio utilizaron los datos procedentes del cribado de anomalías genéticas aprovechando que disponían de los datos necesarios ya que ambos cribados utilizan factores biofísicos y bioquímicos comunes (Park et al., 2013). En otros estudios (Guizani et al. 2017; Di Martino et al., 2019 y Prasad et al., 2021) realizaron el cribado de PE entre las mujeres que acudieron a las pruebas de detección del síndrome de Down durante el primer trimestre. En todos estos casos, el cribado de anomalías genéticas no se ha considerado un riesgo extra de sesgo al ser un cribado poblacional ya establecido en la mayoría de los países y llevarse a cabo de manera rutinaria en el primer trimestre de embarazo.

Algunos aspectos de este dominio podrían afectar más que a la validez interna de los estudios, a la validez externa de los mismos, es decir, a la generalización o extrapolación de sus resultados. Entre estos aspectos se encontraron que la participación en todos los estudios fue voluntaria (sesgo del voluntario sano) y se desconoce si la población que rechazó el cribado presentaba las mismas características que la población voluntaria, además, de todos los embarazos elegibles o reclutados solo analizaron aquellos de los que se disponía información suficiente, perdiéndose aquellos embarazos a los que le faltaban los datos de algunas de las pruebas requeridas para aplicar el algoritmo de riesgo o no se disponía de información sobre el parto. Sin embargo, en todos los estudios la población analizada representaba más del 80% de la población que participó en el cribado y en la mayoría de ellos se analizó más del 90% de la población cribada o elegible.

En general, se ha considerado que la mayoría de los estudios presentaban un riesgo de sesgo de población bajo. Los estudios que muestran un riesgo incierto es por no describir los criterios de inclusión y exclusión (Park et al., 2013; Oliveira et al., 2014 y Mosimann et al., 2017) o no presentar el flujo de pacientes cribados ni indicar las causas de exclusión de participantes (Chaemsaithong et al., 2019 y Chen et al., 2021). El estudio de O' Gorman et al., 2017a no mencionaba estos aspectos, pero hace referencia a un artículo previo donde se

describen y de donde se ha obtenido dicha información (O' Gorman, Wright, Poon, Rolnik, Syngelaki, Wright et al., 2017b).

Sesgo en el dominio de la prueba de estudio (Algoritmo FMF)

El hecho de conocer en muchos casos los resultados de la prueba de referencia (desenlace del embarazo) al analizar los resultados de la prueba a valorar (algoritmo de la FMF) podría suponer un sesgo a la hora de su análisis, por lo que sería recomendable el cegamiento de los investigadores que recogen los resultados del embarazo y de los que realizan la aplicación el algoritmo. Así, en estudios como el de Tan et al., 2018 el riesgo calculado utilizando los métodos basados en el teorema de Bayes no se puso a disposición de los participantes ni de sus médicos y los datos se proporcionaron al estadístico del estudio cegados al resultado del parto. Sin embargo, no todos los estudios mencionan el cegamiento. En este caso al tratarse de un algoritmo matemático con resultado numérico que no implicaba una interpretación subjetiva por parte del investigador se ha considerado que el riesgo de sesgo sería bajo en todos los casos, ya que en todos ellos se indicaba el punto de corte utilizado para las estimaciones.

Figura 3. Valoración del control de sesgos de los estudios incluidos en la revisión



Sesgo en el dominio de prueba de referencia

La prueba de referencia en todos los estudios es la aparición y diagnóstico de preeclampsia, ya que es precisamente esta condición la que intenta detectar la prueba de estudio (algoritmo de la FMF) para poder seleccionar aquellos embarazos candidatos a recibir tratamiento preventivo con aspirina. Todos los estudios utilizaron criterios objetivos y estandarizados para el diagnóstico de la PE, aunque, estos criterios no fueron los mismos en todos los estudios. La mayoría siguieron los criterios establecidos por la International Society for the Study of Hypertension in Pregnancy (ISSHP) y The American College of Obstetricians and Gynecologists (ACOG); si bien algunos utilizaron los criterios establecidos por sociedades e instituciones de ámbito nacional. También, señalar que algunos estudios usaron los criterios tradicionales de la ISSHP 2001 y ACOG 2002 mientras otros estudios usaron los criterios de esas mismas asociaciones revisados y actualizados (Brown et al., 2018, ACOG 2018). De todas formas, se ha considerado que el riesgo es bajo al especificar todos los criterios utilizados. Los resultados perinatales de las cohortes fueron obtenidos de los registros de los centros sanitarios.

Sesgo en el dominio de momento de realización y flujo de pacientes

Por último, con respecto al momento y flujo de pacientes, las pruebas se realizaron en los momentos adecuados no pudiéndose realizar ambas en un corto periodo de tiempo ya que era necesario completar el seguimiento del embarazo para conocer si había PE o no. La mayoría de los estudios analizaron todos los pacientes incluidos en el estudio ya que se habían seleccionado previamente aquellos de los que se disponía la información completa.

Descripción de los estudios

En general, los estudios incluían poblaciones que iban desde 541 a 35.948 mujeres todas ellas reclutadas de manera prospectiva en visitas rutinarias durante el primer trimestre de gestación. A todas las participantes se les realizaba la misma prueba de cribado (algoritmo FMF), aunque no todos los estudios utilizaban el algoritmo con los mismos factores de riesgo y los estudios recogían los mismos datos sobre el resultado del embarazo (identificaban los casos de preeclampsia aparecidos). Las características generales de cada uno de los estudios y la descripción de sus poblaciones según los factores maternos del algoritmo de la FMF se resumen en los Anexos 3 y 4, respectivamente.

Los 16 estudios incluidos analizan poblaciones procedentes de 18 países diferentes. Dos estudios incluían poblaciones norteamericanas (Estados Unidos y Canadá), siete estudios recogían población europea de varios países, un estudio con población sudamericana (Brasil), cinco con población asiática y uno con población australiana. De esta manera, por etnia se distingue un estudio con población mayoritariamente de etnia mixta (Lobo *et al.*, 2019), cinco estudios de etnia asiática (4 del este de Asia: Chaemsaithong *et al.*, 2019; Goto *et al.*, 2021; Chen *et al.*, 2021; Hu *et al.*, 2021 y uno del sur de Asia: Prasad *et al.*, 2021), uno con un 50% de etnia negra (Oliveira *et al.*, 2014) y los restantes 9 estudios con poblaciones mayoritariamente de etnia blanca, de los cuales siete eran con poblaciones europeas. Básicamente los criterios de inclusión en el cribado consistieron en mujeres con embarazos de 11 a 13 semanas de gestación que querían participar en el cribado y daban su consentimiento; y se excluyeron aquellos casos donde no dieron el consentimiento para participar en el estudio (por cualquier razón) y casos de enfermedad materna grave o anomalías fetales detectadas en la revisión rutinaria del embarazo. Finalmente, de las

mujeres que consintieron participar en el cribado de PE solamente se incluyeron en el análisis los embarazos de los que disponían al final del embarazo toda la información necesaria para aplicar el algoritmo de la FMF, también excluyeron de análisis los abortos espontáneos e interrupciones del embarazo que ocurrieron antes de la semana 22 o 24 de gestación. Todos los estudios describían las características de su población por grupos, casos de PE y no afectadas con PE. En el Anexo 4 se muestran los cálculos realizados para la población global de los estudios.

No todos los estudios utilizaban los mismos factores de riesgo incluidos en el algoritmo. Así, por ejemplo, el estudio de Di Martino *et al.*, 2019 calculó el riesgo utilizando solamente las características maternas del algoritmo de la FM, lo que se ha denominado "riesgo *a priori*", mientras que otros estudios han utilizado el denominado triple test (características maternas+ MAP+UtAPI+PIGF) establecido como el de mayor sensibilidad por la FMF o diferentes combinaciones de factores analizando el rendimiento diagnóstico de cada una de ellas.

En cuanto al diagnóstico de PE, 14 estudios utilizaron criterios clínicos de organizaciones como ISSHP y ACOG y dos los criterios establecidos por sus sistemas o instituciones a nivel nacional. En cuanto a la actualización de estos criterios, nueve estudios utilizaron para el diagnóstico de la PE la aparición de hipertensión de *novo* a partir de la semana 20 de gestación acompañada de proteinuria (ISSHP 2001 y ACOG 2002) y siete estudios realizaron el diagnóstico con criterios clínicos más actualizados en los que ya no se consideraba la proteinuria como síntoma obligatorio y se incluyen la aparición de otras condiciones clínicas como otras disfunciones orgánicas materna (insuficiencia renal, compromiso hepático, complicaciones neurológicas, complicaciones hematológicas) y/o disfunción uteroplacentaria. Los criterios utilizados por los estudios que utilizaron los *establecidos a nivel nacional se ajustaban a los criterios actualizados*.

Aunque todos los estudios realizaron cribados prospectivos de embarazos no seleccionados en consultas rutinarias de control, tres estudios mostraban una prevalencia de PE superior a la de los demás, con una PE global entre 4,87% y 5,51%, mientras que en el resto de estudios la prevalencia oscila entre 0,89% y 3,88%. De los estudios con mayor prevalencia: Boutin *et al.*, 2021 incluía una población exclusivamente de mujeres nulíparas, los otros dos estudios Lobo *et al.*, 2019 y Oliveira *et al.*, 2014 se llevaron a cabo en Brasil y Estados Unidos e incluían poblaciones mayoritariamente de etnia mixta y de etnia negra, respectivamente. En estos estudios las poblaciones podrían afectar a la prevalencia ya que, por ejemplo, en la etnia negra la prevalencia de PE es mayor (Zhang, Wan, Ng, Singh, Cheng, Velickovic *et al.*, 2020). Las prevalencias de PE tanto global como la temprana y pretérmino encontradas en cada estudio se ha resumido en la Tabla 2.

Tabla 2. Prevalencia de PE encontrada en los estudios incluidos en la revisión				
Estudio	PE temprana n/N (%)	PE pretérmino n/N (%)	PE total n/N (%)	
Park 2013	12/3014 (0,40%)	-	83/3014 (2,75%)	
Oliveira 2014	30/2969 (1,01%)	-	159/2969 (5,36%)	
Skrästad 2014	1/541 (0,18%)	5/541 (0,92%)	21/541 (3,88%)	
Mosimann 2017	2/1129 (0,18%)	5/1129 (0,44%)	19/1129 (1,68%)	
Rolnik 2017	-	180/25797 (0,77%)	630/25797 (2,44%)	
O´Gorman 2017a	-	59/8775 (0,67%)	239/8775 (2,7%)	
Guizani 2018	18/3239 (0,56%)	36/3239 (1,11%)	80/3239 (2,47%)	
Tan 2018	-	142/16747 (0,84%)	473/16747 (2,82%)	
Di Martino 2019	67/11632 (0,58%)	-	278/11632 (2,39%)	
Lobo 2019	7/617 (1,13%)	18/617 (2,92%)	34/617 (5,51%)	
Chaemsaithong 2019	25/10935 (0,23%)	73/10935 (0,67%)	224/10935 (2,05%)	
Boutin 2021	-	29/4575 (0,63%)	223/4575 (4,87 %)	
Goto 2021	-	11/913 (1,20%)	26/913 (2,84%)	
Chen 2021	-	8/700 (1,14%)	25/700 (3,57%)	
Hu 2021	-	117/10899 (1,07%)	312/10899 (2,86%)	
Prasad 2021	-	25/1863 (1,34%)	59/1863 (3,17%)	

PE temprana: preeclampsia que conduce al parto antes de las 34 semanas de gestación.

PE pretérmino: preeclampsia que conduce al parto antes de las 37 semanas de gestación.

Resultados de rendimiento global de la prueba

Curvas ROC y áreas bajo la curva

No todos los estudios han analizado el rendimiento del algoritmo estimando el área bajo la curva (AUC). En las Tablas 3 y 4 se resume las áreas bajo la curva descritas con las características maternas y con éstas junto a factores biofísicos y factores bioquímicos. Estas comparaciones se realizaron en los estudios para mostrar la variación que implicaba en el rendimiento del algoritmo la incorporación de estos últimos factores en la predicción de PE temprana (< 32-34 semanas) y PE pretérmino (< 37 semanas), respectivamente.

De los 19 estudios solamente seis aportaban el rendimiento del algoritmo para la detección del riesgo de PE temprana (Tabla 3).

Tabla 3. Rendimiento global del modelo de predicción de PE temprana de la FMF **Estudio AUC (IC95%)** CM Triple test Cuádruple test Park 2013 0,755 (0,739-0,771) 0,926 (0,916-0,936)* Oliveira 2014 0,80 (0,71-0,89)* O'Gorman 2017a, b 0,9870/0,9644* 0,8045 0,9865 Guizani 2018 0,772 (0,757-0,786) 0,913 (0,903-0,923)* 0,937 (0,928-0,945) Di Martino 2019 0,764 (0,692-0,836) Lobo 2019 0,946 (0,919-0,973) CM: características maternas

Triple test: CM+MAP+UtAPI+PIGF /* con PAPP-A en lugar de PIGF

Cuádruple test: CM+MAP+UtAPI+PIGF + PAPP-A

Tabla 4. Rendimiento global del modelo de predicción de PE pretérmino de la FMF					
Estudio	AUC (IC95%)				
	СМ	Triple test	Cuádruple test		
Park 2013	0,755 (0,739-0,771)	0,926 (0,916-0,936)*	-		
Oliveira 2014	-	0,80 (0,71-0,89)*	-		
Skrästad 2014	-	-	0,936 (0,860-1,000)		
Mosimann 2017	-	-	-		
Rolnik 2017	-	-	-		
O´Gorman 2017a, b	0,7583	0,9242/0,8956*	0,9241		
Guizani 2018	0,790 (0,776-0,804)	0,910 (0,900-0,920)*	0,932 (0,923-0,940)		
Tan 2018	-	-	-		
Di Martino 2019	0,764 (0,692-0,836)	-	-		
Lobo 2019	-	-	0,870 (0,798-0,942)		
Chaemsaithong 2019	0,758 (0,749-0,766)	0,857 (0,851-0,864)	-		
Boutin 2021	-	-	-		
Goto 2021	-	0,948 (0,863-0,981)	-		
Chen 2021	-	0,7876 (0,648-0,927)	-		
Hu 2021	0,7365 (0,689-0,784)	0,8347 (0,794-0,875) / 0,8600 (0,823-0,897)*	0,843 (0,804-0,882)		
Prasad 2021	0,76 (0,63-0,89)	-	0,96 (0,95-0,98)		

CM: características maternas

Triple test: CM+MAP+UtAPI+PIGF /* con PAPP-A en lugar de PIGF Cuádruple test: CM+MAP+UtAPI+PIGF + PAPP-A

El AUC adopta valores entre 0 y 1 (0% y 100%), delimitando un cuadrado de área igual a 1,00. Un test diagnóstico se considera no-discriminativo si su AUC = 0,50 y a medida que el AUC de un test diagnóstico se acerca al valor 1,00 (test diagnóstico perfecto), mayor será su capacidad discriminativa. Como se observa en las tablas 3 y 4, las 3 combinaciones de factores analizadas presentaban AUC superiores a 0,5 lo que indicaba que presentaban capacidad de discriminación diagnóstica.

El modelo de predicción de la FMF para la PE temprana, mostraba que las AUC se incrementaban al incorporar los factores biofísicos y bioquímicos; pasando de 0,755-0,77 con las características maternas, a 0,80-0,98 cuando se realizaba el triple o cuádruple test. Los intervalos de confianza al 95% (IC95%) no incluían el valor 0,5 lo que indicaba que en todos los casos había discriminación entre riesgo alto y bajo; además, no se observaba solapamiento entre los intervalos de confianza de las AUC con las características maternas y los de los modelos con los factores biofísicos y bioquímicos lo que sugería que el incremento observado sería estadísticamente significativo

En el caso del modelo de la FMF para la PE pretérmino, las AUC observadas cuando se usaban solamente los factores maternos se encontraban en un rango entre 0,736 y 0,790; las observadas para el triple test se encontraban entre 0,787-0,948 y para el cuádruple test 0,843 y 0,96. En todos los casos, tanto el AUC como los límites inferiores de sus IC 95% eran mayores de 0,5 lo que indicaba que la prueba presentaba capacidad de discriminar entre riesgo alto y bajo de riesgo de PE. Además, las áreas obtenidas con los parámetros maternos eran aproximadamente de 0,7 mostrando una capacidad de discriminación regular; sin embargo, las AUC al incorporar a las características maternas los factores físico-químicos aumentaban y se encontraban entre 0,8-0,9 lo que indicaba una buena o muy buena capacidad de discriminación.

En resumen, las AUC aportadas por los estudios han mostrado que la incorporación de los factores bioquímicos y biofísicos al modelo incrementaba el rendimiento del mismo. El triple test presentaba una buena capacidad de discriminación y era similar a la del cuádruple test.

Tasas de detección

La mayoría de los estudios han estimado la tasa de detección de la prueba. Las tasas de detección permiten valorar la probabilidad individual del riesgo de PE temprana y pretérmino. Es una medida de la sensibilidad, y con ellas se han podido estimar los parámetros como especificidad y los coeficientes de probabilidad o verosimilitud.

Aunque varios estudios aportaban las estimaciones considerando varios puntos de corte de falsos positivos: 5%, 10%, 20%... en este informe se ha considerado únicamente como punto de corte el de 10% de falsos positivos. En las Tablas 5 y 6 se han recogido las tasas de detección o la sensibilidad descritas para la PE temprana y la PE pretérmino en los estudios.

Tabla 5. Tasas de detección(DR) o sensibilidad (S) de PE temprana en los cribados de PE con un 10% FP						
Estudio		DR / S (IC95%) %				
	СМ	Mini test	Triple test	Cuádruple test		
Park 2013	DR:39,5 (10,0-75,5)	DR: 41,7 (15,3-72.2)	DR: 91,7 (61,5- 98,6)*	-		
Oliveira 2014	-	-	S: 52*	-		
O´Gorman 2017a	DR: 53 (28-77)	DR: 76 (50-93)	DR: 100 (80-100) DR: 94 (71-100)*	DR: 100 (80-100)		
Guizani 2018	DR: 55,6 (30,8-78,5)	-	DR: 77,8 (52,4- 93,6)*	DR: 83,3 (58,6-96,4)		
Tan 2018	DR: 48,3 (35,2-61,6)	DR: 65,0 (51,6-76,9)	DR: 90,0 (79,5-96,2) DR: 86,7 (75,4- 94,1)*	DR: 90,0 (79,5-96,2)		
Di Martino 2019	DR: 58,2 (45,5-70,2)	-	-	-		
Lobo 2019	-	-	-	S: 85,7		
Boutin 2021	-	-	-	DR: 70,0		

CM: características maternas Mini test: CM+ MAP+ PAPP-A

Triple test: CM+ MAP+ UtAPI+ PIGF /* con PAPP-A en lugar de PIGF

Cuádruple test: CM+ MAP+ UtAPI+ PIGF + PAPP-A

Tabla 6. Tasas de detección(DR) o sensibilidad (S) de PE pretérmino en los cribados de PE con un 10% FP				
Estudio	DR / S (IC95%) %			
	СМ	Mini test	Triple test	Cuádruple test
Park 2013	DR:39,5 (10,0-75,5)	DR: 41,7 (15,3- 72.2)	DR: 91,7 (61,5-98,6)*	-
Oliveira 2014	-	=	S: 52*	-
Skrästad 2014	-	-		S: 80 (28,4-99.5)
Mosimann 2017	-	=	DR: 80	-
Rolnik 2017	=	=	-	DR: 76,7
O´Gorman 2017a	DR: 41 (28-54)	DR: 49 (36-63)	DR: 75 (62-85)	DR: 80 (67-89)
Guizani 2018	DR: 52,8 (35,5-69,6)	=	DR: 77,8 (60,8-89,9)	DR: 80,6 (64,0-91,8)
Tan 2018	-	DR: 53,5 (45,3- 61,7))	DR: 82,4 (76,1-88,7)	-
Di Martino 2019	DR: 58,2 (45,5-70,2)	-	-	-
Lobo 2019	-	-	-	S: 66,7
Chaemsaithong 2019	DR: 40,2 (29,2-41,7)	-	DR:64.0 (53,3-74,7)	-
Boutin 2021	-	-	-	DR: 55,2
Goto 2021	-	DR: 91,0 †	-	DR:91,0
Chen 2021	-	=	DR:50,0	-
Hu 2021	-	DR: 57,3 (48,3- 66,2)	DR: 56,4 (47,4-65,4) DR: 64,9(56,3-73,6)*	DR: 43,6 (34,6-52,6)
Prasad 2021	DR: 40,0	-	DR: 77,5	DR: 80,0

CM: características maternas

Mini test: CM+ MAP+ PAPP-A/ † PIGF en lugar de PAPP-A

Triple test: CM+ MAP+ UtAPI+ PIGF /* con PAPP-A en lugar de PIGF Cuádruple test: CM+ MAP+ UtAPI+ PIGF + PAPP-A

En general, las tasas de detección descritas aumentaban conforme se sumaban factores de riesgo al algoritmo de la calculadora.

Las tasas de detección del algoritmo de la FMF utilizando únicamente las características maternas se encontraban entre 39%-58% y eran similares para la PE temprana y para la PE pretérmino. La incorporación de la presión arterial (MAP) y el factor bioquímico al algoritmo incrementa la tasa de detección del algoritmo (Mini Test) aunque existe una gran variabilidad entre los estudios.

El Triple Test y Cuádruple Test mostraron unas tasas de detección superiores a las anteriores alcanzando en algunos estudios el 80% para la PE pretérmino y hasta más de un 90% para la PE temprana. Aunque, en los estudios de población asiática y americana las tasas de detección del Triple y Cuádruple Test parecían inferiores a las tasas descritas en los estudios europeos, con unos valores entre 50 y 77%.

En todos los casos las tasas de detección estimadas presentaban unos intervalos de confianza del 95% muy amplios lo que indicaba una falta de precisión en estas estimaciones. En algunos casos el IC95% incluía el 0,5 (50%) valor que indica la no discriminación.

Análisis estadístico

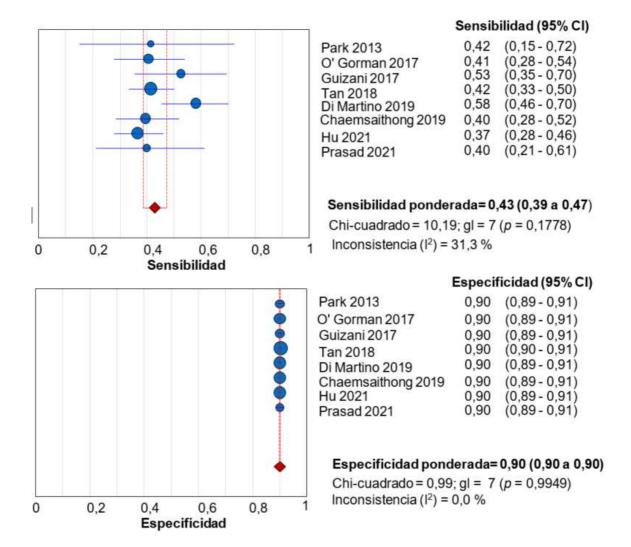
Modelo 1: Validez diagnóstica del algoritmo FMF para el riesgo "a priori" de PE pretérmino.

En la valoración de este modelo se incluyeron los 8 estudios que aportaban información sobre el "riesgo *a priorl*" según el algoritmo de la FMF, con un total de 67.104 embarazos analizados. El análisis estadístico mostró gráficamente la ausencia de efecto umbral en el plano ROC (los puntos en el plano no mostraban un patrón curvilíneo) y se corroboró mediante el coeficiente de Spearman que fue igual a -0,571 (p= 0,139) (Anexo 5).

El análisis estadístico no ha encontrado heterogeneidad entre los estudios y la inconsistencia encontrada era inferior al 40% por lo que se ha utilizado un modelo de efectos fijos para la estimación de los parámetros ponderados. El modelo de la FMF basado únicamente en las características maternas ha mostrado una sensibilidad ponderada muy baja del 43% (IC 95% 0,39-0,47), con un límite superior del IC 95% por debajo de 0,50 por lo que la prueba de este modelo no llega a identificar ni la mitad de los embarazos de alto riesgo de PE.

La especificidad ponderada ha sido muy alta con un 90% y el IC 95% muestra una alta precisión. Por su parte, los CP ponderados con valores de CP+ < de 5 y CP- > de 0,5 mostraban que la prueba de la FMF con características maternas era una prueba de aportación diagnóstica dudosa, lo que ayuda poco a la decisión diagnóstica.

Figura 4. Metanálisis de los valores de sensibilidad y especificidad del modelo de riesgo *a priori*.



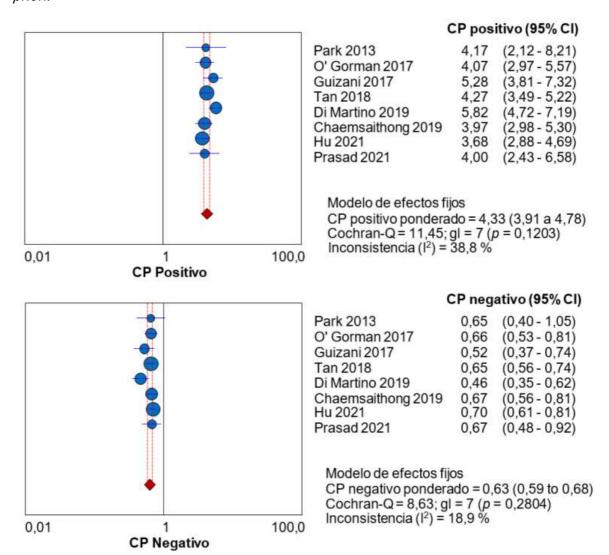


Figura 5. Metanálisis de los coeficientes de probabilidad positivo y negativo del modelo de riesgo *a priori*.

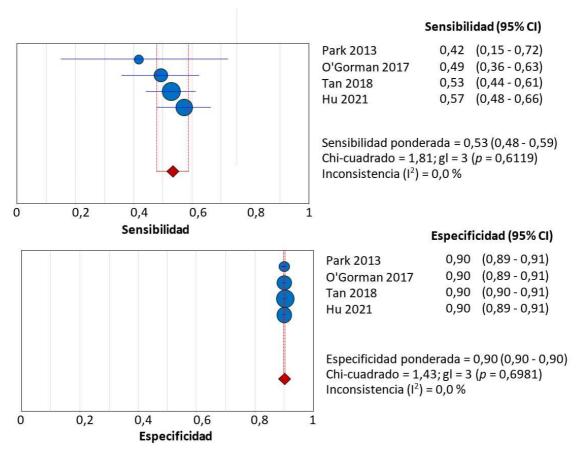
Modelo 2 Validez diagnóstica del Mini test de la FMF para el riesgo de PE pretérmino.

El metanálisis engloba 4 estudios y un total de 39.435 embarazos analizados. El análisis estadístico mostró gráficamente la ausencia de efecto umbral en el plano ROC (los puntos en el plano no mostraban un patrón curvilíneo) y se corroboró mediante el coeficiente de Spearman que fue igual a 0,00 (p = 1,00) (Anexo 6). No se encontró heterogeneidad entre los estudios por lo que se optó por un modelo de efectos fijos.

El Mini Test al igual que la prueba anterior mostraba una especificidad del 90% y una sensibilidad un poco mayor, aunque seguía siendo muy baja, de un 53% con un IC95% que incluía el 50%. Los CP mostraban valores que indicaban que la prueba aportaba algo más de información diagnóstica ya que el CP+ era mayor de 5 y el CP- era más bajo que el anterior,

aunque estos valores siguen siendo propios de una prueba que ayuda poco a la decisión diagnóstica.

Figura 6. Metanálisis de los valores de sensibilidad y especificidad del modelo de mini test.



CP positivo (95% CI) 4.17 (2,12 - 8,21) Park 2013 4,91 O'Gorman 2017 (3,76 - 6,42)Tan 2018 5,48 (4,66 - 6,45) 5.73 (4.85 - 6.76) Hu 2021 Modelo de efectos aleatorios CP positive ponderado= 5,45 (4,90 - 6,05) Cochran-Q = 1,54; gl = 3 (p = 0,6738) Inconsistencia (I2) = 0,0 % 0,01 100,0 CP positivo CP negativo (95% CI) Park 2013 0.65 (0.40 - 1.05) O'Gorman 2017 0.57 (0.44 - 0.73)0,52 (0,44 - 0,62) Tan 2018 Hu 2021 0,47 (0.38 - 0.59)Modelo de efectos aleatorios CP negativo ponderado = 0.52 (0.47 - 0.59)Cochran-Q = 1,97; gl = 3 (p = 0,5776) Inconsistencia (I2) = 0,0 % 0,01 100,0 CP negativo

Figura 7. Metanálisis de los coeficientes de probabilidad positivo y negativo del modelo de mini test.

Modelo 3: Validez diagnóstica del Triple test de la FMF para el riesgo de PE pretérmino.

En el metanálisis se incluyeron 10 estudios con un total de 59.918 mujeres independientemente del tipo de parámetro bioquímico utilizado (PIGF o PAPP-A); sin embargo, tres estudios (O'Gorman 2017, Tan 2018 y Hu 2021) aportaban datos del triple test con PIGF y con PAPP-A de ahí que se hayan incluido en el metanálisis dos veces.

Figura 8. Metanálisis de los valores de sensibilidad y especificidad del modelo de triple test.

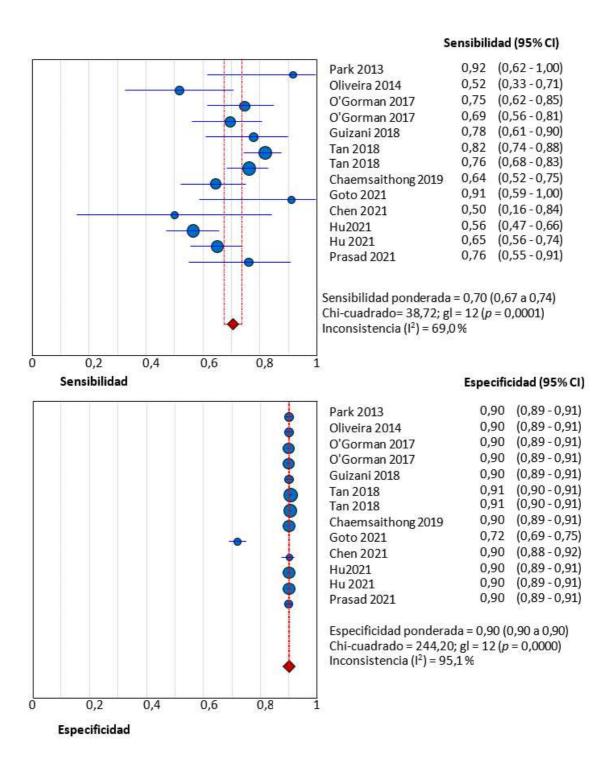
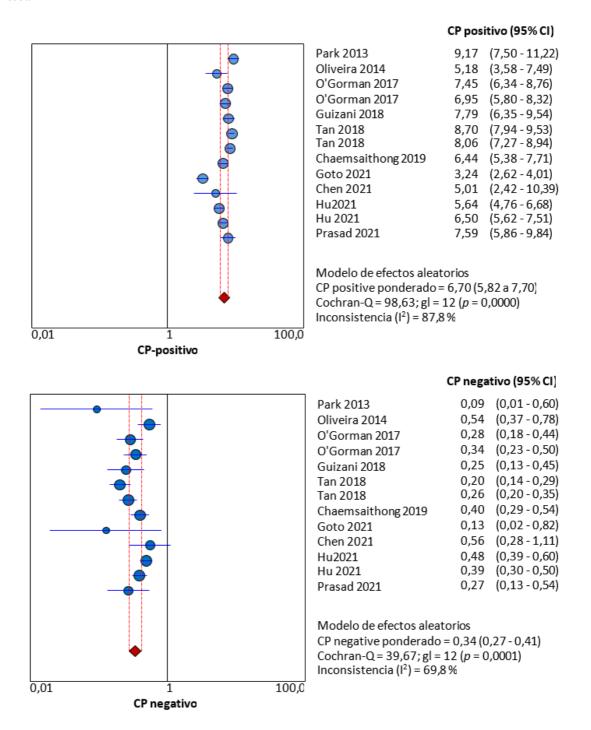


Figura 9. Metanálisis de los coeficientes de probabilidad positivo y negativo del modelo de triple test.



El análisis estadístico mostró gráficamente la ausencia de efecto umbral en el plano ROC (los puntos en el plano no mostraban un patrón curvilíneo) y se corroboró mediante el coeficiente de Spearman que fue igual a 0,085 (p = 0,781) (Anexo 7).

El Triple Test ha mostrado una sensibilidad del 70% y una especificidad del 90%. Además, el CP+ (> 6) se acercaba al valor de 10 y el CP- disminuía (0,34) y se iba acercando al 0,1, lo que indicaba que la prueba era mejor que las anteriores.

Estos datos indicaban que la incorporación de los factores biofísicos (MAP y UtAPI) junto con un factor bioquímico (PIGF o PAPP-A) aumentaba mucho el rendimiento de la prueba. Sin embargo, se encontró heterogeneidad e inconsistencia entre los estudios.

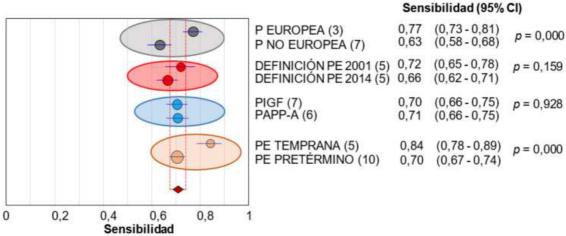
Análisis de sensibilidad

Se analizaron varios subgrupos con el fin de explorar las posibles causas de la heterogeneidad entre los estudios. Así, en función del origen de la población se analizaron dos subgrupos: población europea y población no europea. En función del parámetro bioquímico utilizado en el triple test se valoraron también dos grupos: triple test con PAPP-A y triple test con PIGF y en función de la definición de PE utilizada se valoraron: Definición PE según criterios antiguos (2001) y según criterios actualizados (2014) (Anexo 8).

Se ha observado una diferencia en la sensibilidad del algoritmo entre poblaciones europeas y no europeas, 77% (IC95% 73-81%) ν s 63% (IC 95% 58-68%), y las diferencias observadas eran significativas (p=0,000). Por otra parte, ni la definición de preeclampsia usada en su diagnóstico (2001 ν s 2014) ni el factor bioquímico usado en el algoritmo (PIGF vs PAPP-A) mostraron diferencias significativas (p=0,159 y p = 928, respectivamente). En resumen, el análisis de sensibilidad ha mostrado que el Triple Test presentaba mayor rendimiento en población europea (mayoritariamente de etnia blanca) que en población no europea (mayoritariamente de etnia asiática, negra o mixta). En la figura 8 se resume el análisis de subgrupos.

Por último, algunos de los estudios que valoraron el triple test de la FMF aportaron las tasas de detección tanto para la PE pretérmino como para la PE temprana, por lo que aprovechando estos datos se ha analizado el rendimiento de esta prueba en la detección del riesgo de ambas PE, en términos de sensibilidad. Este análisis ha mostrado que la prueba presentaba mayor rendimiento en la detección del riesgo de preeclampsia temprana que de preeclampsia pretérmino, con unas sensibilidades del 84% (IC 95% 78-89%) y 70% (IC 95% 67-74%) siendo esta diferencia estadísticamente significativa (p=0,000). (Anexo 9).

Figura 10. Análisis de sensibilidad y comparación de sensibilidades entre PE temprana y PE pretérmino del modelo del triple test.



Modelo 4: Validez diagnóstica del Cuádruple test de la FMF para el riesgo de PE pretérmino.

El metanálisis engloba 10 estudios y un total de 74.182 embarazos. No se incluyó el estudio de Cordisco $et\,al.$, 2021 ya que al tratarse de un estudio de efectividad del cribado los embarazos identificados de alto riesgo fueron tratados con aspirina lo que puede afectar a la prevalencia de PE pretérmino y comprometer el rendimiento de la prueba El análisis estadístico mostró gráficamente la ausencia de efecto umbral en el plano ROC (los puntos en el plano no mostraban un patrón curvilíneo) y se corroboró mediante el coeficiente de Spearman que fue igual a -0,399 (p= 0,254) (Anexo 10). Se encontró heterogeneidad entre los estudios por lo que se optó por un modelo de efectos aleatorios.

Sensibilidad (95% CI) 0,80 (0,28 - 0,99) Skrastad 2014 0.80 (0.28 - 0.99)Mosimann 2017 0,77 (0.70 - 0.83)Rolnik 2017 0,80 (0,67 - 0,89)O'Gorman 2017 0,81 (0,64 - 0,92)Guizani 2018 0,81 (0.74 - 0.87)Tan 2018 0,67 (0,41 - 0,87)Lobo 2019 0,55 (0,36 - 0,74)Boutin 2021 0.61 (0.51 - 0.70)Hu 2021 0,80 (0.59 - 0.93)Prasad 2021 Sensibilidad ponderada = 0.74 (0.70 a 0.77) Chi-cuadrado = 22,23; gl = 9 (p = 0,0082)Inconsistencia (I2) = 59,5 % 0 0,2 0.4 1 0.6 8,0 Sensibilidad Especificidad (95% CI) 0.90 (0.87 - 0.92) Skrastad 2014 0.89 (0.87 - 0.91)Mosimann 2017 0.90 (0.90 - 0.90)Rolnik 2017 0,90 (0.89 - 0.91)O'Gorman 2017 0.90 (0.89 - 0.91)Guizani 2018 0,91 (0.90 - 0.91)Tan 2018 0,90 (0.87 - 0.92)Lobo 2019 0,89 (0.89 - 0.90)Boutin 2021 0,90 (0.89 - 0.91)Hu 2021 0,90 (0.89 - 0.91)Prasad 2021 Especificidad ponderada = 0,90 (0,90 a 0,90) Chi-cuadrado = 8,39; gl = 9 (p = 04955)

0

0,2

0,4

0,6

Especificidad

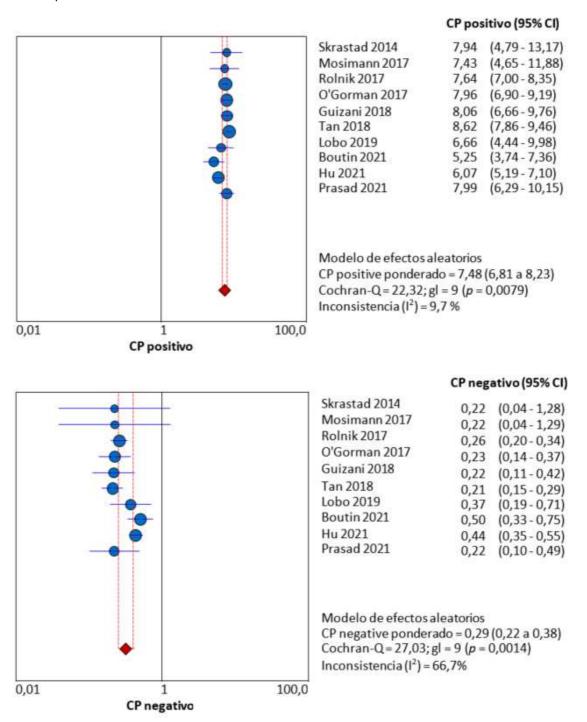
0.8

1

Inconsistency (I-square) = 0,0 %

Figura 11. Metaanálisis de los valores de sensibilidad y especificidad del modelo de Cuádruple Test.

Figura 12. Metaanálisis de los coeficientes de probabilidad positivo y negativo del modelo de Cuádruple Test.



El Cuádruple Test mostraba una sensibilidad y especificidad similar al Triple Test, 74% y 90% respectivamente y valores de CP un poco mejores; sin embargo, estas diferencias entre los dos test no fueron estadísticamente significativas (p= 0,138).

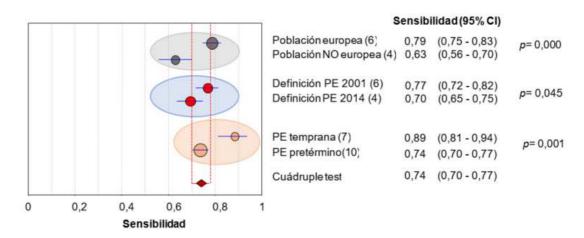
Análisis de sensibilidad

Se analizaron varios subgrupos con el fin de explorar las posibles causas de la heterogeneidad. En función del origen de la población se analizaron dos subgrupos: población europea y población no europea. En función de la definición de PE utilizada en los estudios se valoraron también dos grupos: PE según criterios antiguos (2001) y según criterios actualizados (2014) (Anexo 11).

En la Figura 13 se resume el análisis de subgrupos. Se observa una diferencia en la sensibilidad del algoritmo entre poblaciones europeas y no europeas, 79% (IC 95% 75-83%) νs 63% (IC 95% 56-70%). No se observa solapamiento entre los intervalos de confianza y las diferencias observadas son significativas (p = 0,000). Por su parte, la definición de preeclampsia usada en su diagnóstico, aunque también muestra diferencias en sensibilidad esta diferencia roza la no significación estadística (p=0,045) y sus intervalos de confianza se solapan.

Por último, algunos de los estudios que valoraron el test cuádruple de la FMF aportaron las tasas de detección tanto para la PE pretérmino como para la PE temprana, por lo que aprovechando estos datos se ha analizado el rendimiento de esta prueba en la detección del riesgo de ambas PE en términos de sensibilidad. Este análisis ha mostrado que la prueba presentaba mayor rendimiento en la detección del riesgo de preeclampsia temprana que de preeclampsia pretérmino, con unas sensibilidades del 89% (IC 95% 81-94%) y 74% (IC 95% 70-77%) siendo esta diferencia estadísticamente significativa (p = 0,001). (Anexo 12).

Figura 13. Análisis de sensibilidad y comparación de sensibilidades entre PE temprana y PE pretérmino del modelo de Cuádruple Test.



En las tablas 7 y 8 se resumen los estadísticos ponderados de validez diagnóstica de los cuatro modelos analizados del algoritmo de la FMF para el riesgo de PE pretérmino.

Tabla 7. Resumen	de los estimadores	ponderados d	le sensibilidad,	especificidad,	CP+ y CP- de los
diferentes modelos	del algoritmo de la	FMF nara PF ni	retérmino		

		C (IC OFO)		CD : (IC 0F0/)	CD (IC 050()
Modelo	Estudios(n)/N	S (IC 95%)	E (IC 95%)	CP+ (IC 95%)	CP- (IC 95%)
СМ	8/67.104	0,43 (0,39-0,47)	0,90 (0,90-0,90)	4,33 (3,91-4,78)	0,63 (0,59-0,68)
Mini test	4/39.435	0,53 (0,48-0,59)	0,90 (0,90-0,90)	5,45 (4,90-6,05)	0,52 (0,47-0,59)
Triple test	13/96.339*	0,70 (0,67-0,74)	0,90 (0,90-0,90)	6,70 (5,82-7,70)	0,34 (0,27-0,41)
P. Europea		0,77 (0,73-0,81)			
P.NO europea		0,63 (0,58-0,68)			
PE temprana		0,84 (0,78-0,89)			
Cuádruple test	10/74.182	0,74 (0,70-0,77)	0,90 (0,90-0,90)	7,48 (6,81-8,23)	0,29 (0,22-0,38)
P. Europea	6/	0,79 (0,75-0,83)			
P.NO europea	4/	0,63 (0,56-0,70)			
PE temprana	7/	0,89 (0,81-0,94)			

^{*10} estudios diferentes con 59.918 participantes pero 3 de ellos incluidos dos veces por aportar datos del test con dos parámetros bioquímicos diferentes.

Tabla 8. Comparación entre las sensibilidades ponderadas de los 4 modelos del algoritmo de la FMF.

algoritmo de la FMF.			
Modelo	Estudios(n)/N	S (IC95%)	p
СМ	8 / 67.104	0,43 (0,39-0,47)	-
Mini test	4 / 39.435	0,53 (0,48-0,59)	0,003
Triple test	13 / 96.339*	0,70 (0,67-0,74)	0,000
Cuádruple test	10 / 74.182	0,74 (0,70-0,77)	0,138

^{*10} estudios diferentes con 59.918 participantes, pero 3 de ellos incluidos dos veces por aportar datos del test con dos parámetros bioquímicos diferentes.

La p indica la significación estadística entre el modelo y el modelo anterior.

En resumen:

- La incorporación de los factores bioquímicos y biofísicos a las características maternas incrementaba el rendimiento del algoritmo de la FMF para el riesgo de PE pretérmino.
- El Mini Test, aunque mejora de forma significativa la sensibilidad del "riesgo a priori", presentaba un rendimiento diagnóstico muy bajo e insuficiente para la buena identificación de los casos de alto riesgo de PE pretérmino.
- El triple test de la FMF ha mostrado una buena capacidad de discriminación entre alto y bajo riesgo de PE pretérmino, aunque esta capacidad parece variar en función de las características de la población. En población europea ha mostrado buena capacidad de discriminación con una sensibilidad mayor del 75%; sin embargo, en otras poblaciones (asiática y americana) su capacidad es inferior mostrando una sensibilidad del 63%.
- El cuádruple test presentaba una capacidad de discriminación similar a la del triple test.
- El triple test parece identificar con mayor precisión el riesgo alto de PE temprana.

Discusión

En el caso de la preeclampsia la importancia del problema de salud, por las graves consecuencias que conlleva a la madre y al feto, y la existencia de un tratamiento preventivo, eficaz y seguro, han llevado a varias organizaciones como la US Preventive Services Task Force (USPSTF), The International Society for the Study of Hypertension in Pregnancy (ISSHP) o The International Federation of Gynecology and Obstetrics (FIGO) a recomendar el cribado de PE. La USPSTF no encontró evidencia adecuada para recomendar ninguna prueba de predicción de riesgo como, por ejemplo, indicadores clínicos, marcadores séricos o el UtAPI como estrategia de detección de PE, y solo ha considerado que la detección de PE mediante mediciones de la presión arterial tiene un beneficio neto sustancial (USPSTF, 2017). Por otro lado, la ISSHP reconoce que ninguna prueba o conjunto de pruebas del primer o segundo trimestre puede predecir de manera confiable el desarrollo de todos los casos de preeclampsia; sin embargo, una combinación de factores de riesgo maternos, presión arterial, factor de crecimiento placentario (PIGF) y Doppler de la arteria uterina puede seleccionar mujeres que pueden beneficiarse de 150 mg/día de aspirina para prevenir el parto prematuro (antes de las 37 semanas de gestación) pero no la preeclampsia a término. A pesar de ello, la ISSHP apoya la detección del riesgo de preeclampsia en el primer trimestre cuando esto se puede integrar en el sistema de salud local, aunque aún no se ha establecido la rentabilidad de este enfoque. La FIGO en la misma línea recomienda el uso de cribados para el riesgo de PE basados en biomarcadores frente a los métodos que solo utilizan las características demográficas y el historial médico materno (Poon, Shennan, Hyett, Kapur, Hadar, Divakar et al., 2019). Por lo tanto, aunque se recomienda el cribado para la detección precoz de la preeclampsia, existe incertidumbre sobre la prueba más adecuada para el cribado.

En esta revisión sistemática se ha valorado el rendimiento diagnóstico del algoritmo de predicción propuesto por la FMF como prueba para el cribado de la PE pretérmino durante el primer trimestre de gestación. El análisis estadístico llevado a cabo en esta revisión sistemática ha mostrado que la incorporación de los factores bioquímicos y biofísicos a las características maternas incrementaba de manera significativa la identificación de los casos de embarazos de alto riesgo de PE pretérmino. Se ha observado que la incorporación de manera progresiva de dichos parámetros incrementaba la sensibilidad del test sin afectar a la especificidad del mismo, al mismo tiempo que aumentaba el CP+ y disminuía el CP-. En el análisis se han valorado los CP y no los valores predictivos, ya que estos últimos dependen de la prevalencia de la enfermedad y en los casos como el de la preeclampsia en el que la prevalencia es baja hasta las pruebas excelentes tienen un valor predictivo positivo bajo. Por lo tanto, se ha preferido estimar los CP ya que los valores predictivos relacionan las características de la prueba con las poblaciones, mientras que los cocientes de probabilidad se pueden aplicar a un paciente específico.

De las posibles combinaciones de factores de riesgo disponibles en la calculadora de la FMF la que presentaba mayor rendimiento con el mínimo de factores de riesgo era el denominado Triple Test, la incorporación de un solo factor bioquímico mostraba un rendimiento similar al obtenido cuando se incorporaban los dos, independientemente del factor utilizado PAPP-A o PIGF. A pesar de ser el de mayor rendimiento, la sensibilidad ponderada del 74% mostrada por el Triple Test no es tan alta como se esperaría para una prueba de cribado dónde al menos se desearía una sensibilidad mayor del 80%. Además, el

Triple Test presentaba un CP+ ponderado de 6,70 (IC 95% 5,82 a 7,70) y un CP- ponderado de 0,34 (IC 95% 0,27-0,41), si tenemos en cuenta que los CP+ que son de utilidad para apoyar el diagnóstico han de ser mayores de 5, idealmente mayores de 10 y que los CP- útiles para descartar la condición son los menores de 0,2, idealmente 0,1; el Triple Test podría ser considerada una prueba que ayuda a la identificación del riesgo alto de PE pero no ayuda a descartarlo. A pesar de que el Triple Test no pueda considerarse una prueba excelente, según estos parámetros ponderados, parece mejor que el método actual de identificación del riesgo de PE pretérmino basado en las características maternas, este último con una sensibilidad ponderada del 43%, un CP+ ponderado < de 5 y un CP- ponderado > de 0,5 con una ayuda nula o dudosa a la decisión diagnóstica. Por otra parte, la sensibilidad del Triple Test aumentaba hasta un 84 % para la identificación de alto riesgo de PE temprana, con un CP+ ponderado > 8 y un CP- ponderado < 0,2. Esto significaría que de los casos positivos para la PE pretérmino se identificarían prácticamente todos los casos de PE temprana. Esta identificación de la PE temprana es clínicamente muy relevante porque estos casos llevan asociado un alto riesgo de parto muy prematuro o extremadamente prematuro con gran riesgo para la madre y el bebé.

El estudio de subgrupos mostró que ni la definición de PE ni el factor bioquímico utilizado (PAPP-a o PIGF) afectaba el rendimiento del Triple Test. Si bien, el análisis de estos subgrupos mostró heterogeneidad entre los estudios y una inconsistencia moderada. Los resultados obtenidos sobre el efecto de las actualizaciones en los criterios de definición de la PE coinciden con los presentados por Khan *et al.*, 2019, que observaron que la nueva clasificación de PE incrementaba la incidencia de PE alrededor de un 21% con los criterios de ISSHP y un 7% con los de ASCOG mientras que el rendimiento del modelo triple test era similar.

Por el contrario, el origen de la población y en este caso la etnia sí tenía un efecto claro y significativo en el resultado de la prueba. Se ha observado que el rendimiento de la prueba era mayor en población europea, mayoritariamente blanca, que en otras poblaciones en las que las etnias predominantes no eran la blanca. No hay que olvidar que un algoritmo no es ni más ni menos que una estimación matemática de un riesgo en función de una serie de factores. En concreto el algoritmo de la FMF se desarrolló y se validó inicialmente en una población procedente del Reino Unido, por lo que se ajusta a los resultados del metanálisis realizado ya que en nuestro análisis de subgrupos los estudios incluidos con población europea eran solamente 3, dos de ellos realizados en Reino Unido y, además, se habían incorporado dos veces para la estimación ponderada al aportar datos con PIGF y con PAPP-A. Las poblaciones con mayoría de otras etnias presentaban un rendimiento inferior y esto puede deberse a que los valores de las MoM de los factores de riesgo de esas poblaciones difirieran de los de la población de validación inicial. En el metanálisis no se ha podido hacer análisis de subgrupos por etnias concretas ya que no se disponía de estudios suficientes. La mayoría de los estudios no europeos eran asiáticos por lo que los resultados obtenidos reflejan principalmente el rendimiento en esta población. Un estudio sobre los perfiles de los biomarcadores en población asiática ha mostrado que debido a las diferencias antropométricas entre las mujeres asiáticas y las caucásicas, aparecen diferencias significativas en los valores MoM de biomarcadores para la detección de PE, en particular en las MoM de los biomarcadores MAP y PIGF. El estudio consideraba que se requería el ajuste de factores adicionales o el desarrollo de fórmulas específicas en Asia para el cálculo de los MoM de biomarcadores (Chaemsaithong et al., 2020). De hecho, los estudios asiáticos incluidos indicaban que habían determinado las MoM de los biomarcadores específicos de Asia

mediante un análisis de regresión multivariante (Chaemsaithong et al., 2019; Chen et al., 2021). A pesar de este ajuste en las MoM de los biomarcadores en el análisis estadístico persiste el menor rendimiento de la prueba en esta población, los propios autores de los estudios sugieren que la diferencia en el rendimiento del Triple Test podría deberse a que tanto la frecuencia de PE prematura y como de factores de riesgo en las mujeres asiáticas fuera menor (Chaemsaithong et al., 2019). Uno de los estudios asiáticos, Goto et al., 2021 llevado a cabo en población japonesa mostró una tasa de detección mayor que la del resto de estudios asiáticos y más cercana a la de los estudios europeos; en este estudio su población presentaba una MoM de PIGF mayor que la de los otros estudios asiáticos. Estos resultados ponen de manifiesto que del Triple Test de la FMF aunque mejora de forma general la predicción del riesgo alto de PE el perfil de la población afecta a su rendimiento. Algunos investigadores opinan que los trabajos de validación utilizando el algoritmo FMF deberían aportar la información detallada de las distribuciones de los factores de riesgo de sus poblaciones con el fin de que puedan servir para analizar las fuentes de discrepancias y permitir así mejorar el rendimiento del modelo y un ajuste para cada población específica (Farina y Cavoretto, 2021).

Entre las fortalezas del análisis estadístico realizado se encuentran el número de estudios incluidos y una gran población analizada, así como la diversidad en su procedencia con representación de poblaciones de 18 países. Los estudios eran similares en cuanto a diseño y con el reclutamiento de mujeres de forma prospectiva, en general con bajo riesgo de sesgo. Todos los metanálisis incluían al menos tres estudios con poblaciones diferentes y ni los cambios en los criterios de diagnóstico de la PE ni el factor bioquímico utilizado afectaba al resultado de los parámetros ponderados. Los estudios han presentado la tasa de detección o la sensibilidad para una especificidad fija (el 10% de falsos positivos). Esto elimina el requisito de tener en cuenta la correlación entre la sensibilidad y la especificidad entre los estudios ya que todas las especificidades tienen el mismo valor.

Entre las limitaciones de esta revisión se encuentran que los resultados aportados por los estudios mostraban en general IC 95% muy amplios, debido a que recogían muy pocos casos de preeclampsia pretérmino (prevalencia muy baja) lo que afectaba a la precisión de los resultados. Además, el Triple Test es un modelo predictivo y, por lo tanto, se trata de una ecuación matemática que describe la relación entre uno o más biomarcadores pronósticos y un resultado dado. Dos problemas fundamentales pueden ocurrir cuando se incluyen biomarcadores en modelos estadísticos para predecir un resultado clínico: 'error en la discriminación' y 'error en calibración'. La discriminación refleja la capacidad del modelo para distinguir un estado (muerto/sobreviviente, evento/no evento), mientras que la calibración mide cuánto coincide la estimación del pronóstico de un modelo predictivo con probabilidad real de resultado (es decir, la proporción observada del evento) (Tripepi, Jager, Dekker y Zoccali, 2010). La calibración se refiere por lo tanto a la concordancia entre los resultados observados y las predicciones. En el caso del modelo Triple Test de la FMF se valoró la discriminación y calibración del modelo en poblaciones del Reino Unido mostrando que presentaba buena discriminación (AUC mayores del 0,8) y calibración (buen acuerdo entre los riesgos predichos por el modelo y la incidencia observada) (Wright et al., 2019). Dado que el modelo presentaba diferencias en su rendimiento diagnóstico habría sido importante estudiar la calibración del modelo en las distintas poblaciones para la validación externa del Triple Test. Sin embargo, es importante resaltar que los estudios incluidos en este análisis estadístico se han centrado en la valoración de la capacidad de discriminación de la prueba analizando el AUC y las tasas de detección y solamente dos estudios con población asiática aportaban la calibración del modelo en su población (Chaemsaithong *et al.*, 2019; Goto et al., 2021). En estos estudios los resultados de la calibración fueron contradictorios, el estudio de Chaemsaithong *et al.*, 2019 encontró que el modelo Triple Test estaba bien calibrado mientras que en el estudio de Goto *et al.*, 2021 el modelo no mostraba una buena calibración. Estas diferencias podrían deberse al tamaño de muestra de Goto *et al.*, 2021 que incluía solo 913 embarazos y al bajo número de casos de PE pretérmino observados (11 casos); el estudio de Chaemsaithong *et al.*, 2019 incluía una muestra de 10.935 embarazos con un número de casos observados de 73. Pero, no ha sido posible valorar la calibración del modelo en diferentes poblaciones.

El modelo propuesto por la FMF no es el único con una calculadora *online* disponible para la estimación del riesgo de PE. En España el centro de Medicina Fetal Barcelona (MFB), ha desarrollado una calculadora *online* similar a la del FMF basada en los mismos factores de riesgo, aunque utilizando un modelo o algoritmo diferente. El MFB es un centro de investigación multidisciplinar en medicina fetal y perinatal vinculado a BCNatal y a la Universidad de Barcelona. BCNatal integra funcionalmente los departamentos de medicina materno-fetal y obstetricia, Ginecología y Neonatología de dos hospitales universitarios: el Hospital Sant Joan de Déu Barcelona, y el Hospital Clínic Barcelona. Esta calculadora utiliza un algoritmo con las características maternas, presión arterial, UtAPI y PAPP-A (Scazzochio *et al.*, 2013). El modelo estimaba una DR de 80,8% para PE temprana (< 34 semanas) y de 39,6% para PE tardía (> 34 semanas) considerando un 10% de FP. A diferencia del modelo de la FMF este modelo no ha sido estudiado en otras poblaciones.

Los estudios que han comparado el rendimiento de la prueba de la FMF con otros algoritmos o criterios de cribado como los del NICE o la ACOG han mostrado que el rendimiento del Triple Test era superior (O' Gorman *et al.*, 2017a); con una tasa de detección del Triple Test del 82,4% frente a la del NICE de 40,8%, (Tan *et al.*, 2018). Pero incluso, se ha observado que en la estimación del "riesgo a priori" (considerando solamente las características maternas) el algoritmo de la FMF presentaba una mayor tasa de detección que el algoritmo de la BCNatal (Scazzochio *et al.*, 2013; Di Martino *et al.*, 2021).

El cribado con la calculadora de la FMF en el primer trimestre de gestación ha sido implementado en Italia y en Reino Unido mostrando su factibilidad y su efectividad (Cordisco et al., 2021; Guy et al., 2021, respectivamente). Así, en Reino Unido la implementación del programa de cribado con el Triple Test de la FMF, en sustitución del que se realizaba previamente con los criterios NICE, mostró una aparente reducción del 23% en la prevalencia de preeclampsia. También redujo de manera significativa la tasa de cribado positivo de PE (8,2% *vs* 16,1%) y las demás medidas de rendimiento como la sensibilidad, la especificidad y los coeficientes de probabilidad. Aunque la AUC de 0,846 demostró una buena discriminación de la prueba para la PE pretérmino. Esta reducción en el rendimiento de cribado se debería al efecto de la aspirina, ya que a los casos identificados como de alto riesgo se trataron con 150 mg de aspirina como tratamiento preventivo. La implementación también mostró que el 99% de los casos de alto riesgo según la FMF cumplieron el tratamiento frente al 28,9% de los casos de cribado positivo según el NICE (Guy et al., 2021). Los cálculos ajustados por el efecto de la aspirina mostrarían que la tasa de detección (sensibilidad) sería del 76% en comparación con la del 55,6% encontrada en el cribado (Wright y Nicolaides, 2021). Estos valores de sensibilidad ajustados están más próximos a los valores descritos en los estudios de validación del algoritmo en Reino Unido y al obtenido en nuestro análisis estadístico.

Además, de estos datos sobre la factibilidad y efectividad del cribado con el algoritmo de la FMF; estudios económicos realizados en Europa (Dubon Garcia *et al.*, 2021), en Canadá

(Ortved, Hawkins, Johnson, Hyett y Metcalfe, 2019) y en Australia (Park, Deeming, Bennett y Hyett, 2021) coinciden en mostrar que el cribado con la FMF y el subsiguiente tratamiento con aspirina en dosis bajas tienen el potencial de prevenir más casos de PE pretérmino que las estrategias basadas en las características maternas en uso actualmente, con un ahorro de costes.

Sin embargo, desde el punto de vista económico incluso se ha valorado el uso universal de aspirina en bajas dosis en mujeres embarazas frente a diferentes estrategias de uso de la aspirina. Los estudios que lo han valorado mostraron que el tratamiento rutinario de aspirina en bajas dosis es más coste-efectivo que cualquier otra estrategia de las estudiadas y por lo tanto sugieren que este sería el enfoque más racional y clínicamente pragmático (Werner, Hauspurg y Rouse, 2015; Mone *et al.*, 2018; Mallampati, Grobman, Rouse y Werner, 2019).

La valoración en concreto del uso rutinario de aspirina frente al cribado con el algoritmo de la FMF se basa por una parte en la imperfecta sensibilidad de la prueba de la FMF y por otra, en el bajo coste de la aspirina, desde la perspectiva de los costos, afecta poco que muchas de las mujeres a las que se les recetó aspirina no se beneficien de ella. Por el contrario, las pruebas del cribado son relativamente caras. Esta valoración no apoyaba la estrategia de FMF y encontraba que la aspirina de rutina era un enfoque más eficaz y menos costoso para la prevención de la preeclampsia en mujeres de bajo riesgo. Además, análisis complementarios mostraban que, si se aplicaba la estrategia de la detección y el tratamiento, el modo de prueba menos rentable era usar la detección de MAP solamente (Mone *et al.*, 2018).

Aunque el tratamiento universal con aspirina se postule en varios estudios como el más coste-efectivo para la prevención de la PE frente a otras estrategias existen aspectos que deben tenerse en consideración antes de su puesta en práctica a nivel clínico:

- el efecto beneficioso de la aspirina en la prevención de la preeclampsia prematura parece depender de la adhesión al tratamiento (Wright et al., 2017b; Cuckle, 2020; Mallampati y Werner, 2020), en este aspecto hay que pensar en que no solo influyen las estrategias planteadas, sino que también el comportamiento y el uso de medicamentos durante el embarazo ya que la adhesión suele ser mayor en mujeres con alto riesgo (Cuckle, 2020; Mallampati y Werner, 2020).
- los efectos adversos asociados a la aspirina. Aunque se considera segura, la aspirina probablemente aumenta levemente la hemorragia posparto > 500 ml (RR 1,06, IC 95 %: 1,00 a 1,12) y probablemente aumente marginalmente el riesgo de desprendimiento de placenta (RR 1,21, IC95% 0,95-1,54 (Duley *et al.,* 2019). Pero esta evidencia procede de estudios realizados en mujeres con riesgo de PE, por lo que en estos casos el beneficio del tratamiento es mayor que los efectos adversos observados; sin embargo, el tratamiento universal con aspirina implica la medicación de mujeres con bajo riesgo por lo que se requiere un ensayo controlado aleatorizado con el poder estadístico adecuado para abordar la eficacia y la seguridad de la aspirina universal en dosis bajas en el embarazo de bajo riesgo en comparación con un enfoque de detección.

Conclusiones

- 1. El algoritmo de la FMF para la identificación del riesgo alto de PE pretérmino presenta buen rendimiento diagnóstico.
- 2. La incorporación de los biomarcadores (biofísicos y bioquímicos) a las características maternas incrementa el rendimiento en el diagnóstico del riesgo de PE pretérmino.
- 3. El triple test de la FMF, que incluye las características maternas, MAP, UtAPI y PPAP-A o PIGF, es la prueba con mayor capacidad de discriminación. La incorporación de los dos biomarcadores PAPP-A y PIGF no incrementa la capacidad de discriminación de la prueba.
- 4. El Triple Test de la FMF ha mostrado su capacidad de discriminación en diferentes poblaciones, aunque no ha sido valorada su calibración.
- 5. El rendimiento del Triple Test se ve afectado por las características de la población; así, el rendimiento es significativamente mayor en población europea que en poblaciones procedentes de otras áreas.

Referencias

- ACOG. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. (2013). *Obstetrics and gynecology*, *122*(5), 1122–1131. https://doi.org/10.1097/01.AOG.0000437382.03963.88
- ACOG Committee on Practice Bulletins--Obstetrics (2002). ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia. Number 33, January 2002. *Obstetrics and gynecology*, *99* (1), 159–167. https://doi.org/10.1016/s0029-7844(01)01747-1.
- ACOG Committee Opinion No. 743: Low-Dose Aspirin Use During Pregnancy. (2018). *Obstetrics and gynecology*, *132*(1), e44–e52. https://doi.org/10.1097/AOG.000000000002708
- August, P., Jeyabalan, A. (2021). Preeclampsia: Prevention. In: C.J. Lockwood & V.A. Barss, Eds., *UpToDate*. https://www.uptodate.com/contents/preeclampsia-prevention?search=preclampsia&source=search_result&selectedTitle=7~150&usage_type=default&display_rank=7#H1424609381.
- August, P., Sibai, B.M. (2022). Preeclampsia: Clinical features and diagnosis. In: C.J. Lockwood & V.A. Barss, Eds, *UpToDate*. https://www.uptodate.com/contents/preeclampsia-clinical-features-and-diagnosis?search=preclampsia&source=search_result&selectedTitle=1~150&us age_type=default&display_rank=1#H34.
- Bartsch, E., Medcalf, K. E., Park, A. L., Ray, J. G., & High Risk of Pre-eclampsia Identification Group (2016). Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ* (Clinical research ed.), 353, i1753. https://doi.org/10.1136/bmj.i1753.
- Boutin, A., Guerby, P., Gasse, C., Tapp, S., & Bujold, E. (2021). Pregnancy outcomes in nulliparous women with positive first-trimester preterm preeclampsia screening test: the Great Obstetrical Syndromes cohort study. *American journal of obstetrics and gynecology*, *224*(2), 204.e1–204.e7. https://doi.org/10.1016/j.ajog.2020.08.008.
- Brown, M. A., Magee, L. A., Kenny, L. C., Karumanchi, S. A., McCarthy, F. P., Saito, ··· & International Society for the Study of Hypertension in Pregnancy (ISSHP) (2018). The hypertensive disorders of pregnancy: ISSHP classification, diagnosis & management recommendations for international practice. *Pregnancy hypertension*, *13*, 291–310. https://doi.org/10.1016/j.preghy.2018.05.004
- Brunelli, E., Seidenari, A., Germano, C., Prefumo, F., Cavoretto, P., Di Martino, D., ··· Farina, A. (2020). External validation of a simple risk score based on the ASPRE trial algorithm for preterm pre-eclampsia considering maternal characteristics in nulliparous pregnant women: a multicentre retrospective cohort study.

- BJOG: an international journal of obstetrics and gynaecology,127(10), 1210–1215. https://doi.org/10.1111/1471-0528.16246.
- Bujold, E., Chaillet, N., & Kingdom, J. (2019). Placental growth factor testing for suspected pre-eclampsia. *Lancet*, *393* (10183), 1775–1776. https://doi.org/10.1016/S0140-6736(19)30958-4.
- Chaemsaithong, P., Pooh, R. K., Zheng, M., Ma, R., Chaiyasit, N., Tokunaka, M., ··· Poon, L. C. (2019). Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population. *American journal of obstetrics and gynecology*, *221*(6), 650.e1–650.e16. https://doi.org/10.1016/j.ajog.2019.09.041.
- Chaemsaithong, P., Sahota, D., Pooh, R. K., Zheng, M., Ma, R., Chaiyasit, N., ··· Poon, L. C. (2020). First-trimester pre-eclampsia biomarker profiles in Asian population: multicenter cohort study. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 56*(2), 206–214. https://doi.org/ 10.1002/uog.21905.
- Chaemsaithong, P., Sahota, D. S., & Poon, L. C. (2020). First trimester preeclampsia screening and prediction. *American journal of obstetrics and gynecology*, *226*(2S), S1071–S1097.e2. https://doi.org/10.1016/j.ajog.2020.07.020.
- Chen, Y. T., Lin, T. Y., Cheng, P. J., Chan, K. S., Huang, H. Y., & Shaw, S. W. (2021). Taiwanese new direction in prediction of early pregnancy preeclampsia. *Taiwanese journal of obstetrics & gynecology*, *60*(1), 66–69. https://doi.org/10.1016/j.tjog.2020.11.009
- Cordisco, A., Periti, E., Antoniolli, N., Lozza, V., Conticini, S., Vannucci, G., ··· Pasquini, L. (2021). Clinical implementation of pre-eclampsia screening in the first trimester of pregnancy. *Pregnancy hypertension*, *25*, 34–38. https://doi.org/10.1016/j.preghy.2021.05.010.
- Cuckle H. (2020). Strategies for Prescribing Aspirin to Prevent Preeclampsia: A Cost-Effectiveness Analysis. *Obstetrics and gynecology*, *135*(1), 217. https://doi.org/ 10.1097/AOG.000000000003631.
- Di Martino, D., Masturzo, B., Paracchini, S., Bracco, B., Cavoretto, P., Prefumo, F., ··· Farina, A. (2019). Comparison of two "a priori" risk assessment algorithms for preeclampsia in Italy: a prospective multicenter study. *Archives of gynecology and obstetrics*, 299(6), 1587–1596. https://doi.org/10.1007/s00404-019-05146-4.
- Dubon Garcia, A., Devlieger, R., Redekop, K., Vandeweyer, K., Verlohren, S., & Poon, L. C. (2021). Cost-utility of a first-trimester screening strategy versus the standard of care for nulliparous women to prevent pre-term pre-eclampsia in Belgium. *Pregnancy hypertension*, 25, 219–224. https://doi.org/10.1016/j.preghy.2021.06.012
- Duley, L., Meher, S., Hunter, K. E., Seidler, A. L., & Askie, L. M. (2019). Antiplatelet agents for preventing pre-eclampsia and its complications. *The Cochrane database of*

- *systematic reviews*, *2019*(10), CD004659. https://doi.org/10.1002/14651858.CD004659.pub3.
- Farina, A., & Cavoretto, P. I. (2021). Comment on "Accuracy of the FMF Bayes theorembased model for predicting preeclampsia at 11-13 weeks of gestation in a Japanese population". *Hypertension research: official journal of the Japanese Society of Hypertension*, *44*(6), 720–721. https://doi.org/10.1038/s41440-020-00587-w
- Goto, M., Koide, K., Tokunaka, M., Takita, H., Hamada, S., Nakamura, M., ··· Poon, L. C. (2021). Accuracy of the FMF Bayes theorem-based model for predicting preeclampsia at 11-13 weeks of gestation in a Japanese population. *Hypertension research: official journal of the Japanese Society of Hypertension*, 44(6), 685–691. https://doi.org/10.1038/s41440-020-00571-4.
- Guizani. M.. Valsamis. J.. Dutemever. V.. Kang. X.. Ceccotti. V.. Khalife. J.. · · · Jani. J. C. (2018). First-Trimester Combined Multimarker Prospective Study for the Detection of Pregnancies at a High Risk of Developing Preeclampsia Using the Fetal Medicine Foundation-Algorithm. *Fetal diagnosis and therapy*, *43*(4), 266–273. https://doi.org/10.1159/000477934
- Guy, G. P., Leslie, K., Diaz Gomez, D., Forenc, K., Buck, E., Khalil, A., & Thilaganathan, B. (2021). Implementation of routine first trimester combined screening for preeclampsia: a clinical effectiveness study. *BJOG: an international journal of obstetrics and gynaecology*, *128*(2), 149–156. https://doi.org/ 10.1111/1471-0528.16361
- He, B., Hu, C., & Zhou, Y. (2021). First-trimester screening for fetal growth restriction using Doppler color flow analysis of the uterine artery and serum PAPP-A levels in unselected pregnancies. *The iournal of maternal-fetal & neonatal medicine:* the official iournal of the European Association of Perinatal Medicine. the Federation of Asia and Oceania Perinatal Societies. the International Society of Perinatal Obstetricians, 34(23), 3857–3861. https://doi.org/10.1080/14767058.2019.1701646.
- Hu, J., Gao, J., Liu, J., Meng, H., Hao, N., Song, Y., ··· Sun, Y. (2021). Prospective evaluation of first-trimester screening strategy for preterm pre-eclampsia and its clinical applicability in China. *Ultrasound in obstetrics & gynecology: the official iournal of the International Society of Ultrasound in Obstetrics and Gynecology, 58*(4), 529–539. https://doi.org/10.1002/uog.23645
- Henderson, J.T., Vesco, K.K., Senger, C.A., Thomas, R.G., Redmond, N. (2021). Aspirin Use to Prevent Preeclampsia and Related Morbidity and Mortality: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA 326* (12):1192-1206. https://doi.org/10.1001/jama.2021.8551.
- Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*, *271*(9), 703–707.

- https://doi.org/10.1001/jama.271.9.703
- Lakovschek, I. C., Ulrich, D., Jauk, S., Csapo, B., Kolovetsiou-Kreiner, V., Mayer-Pickel, K., ··· Cervar-Zivkovic, M. (2018). Risk assessment for preterm preeclampsia in first trimester: Comparison of three calculation algorithms. *European journal of obstetrics, gynecology, and reproductive biology, 231*, 241–247. https://doi.org/10.1016/j.ejogrb.2018.11.006.
- Lobo, G., Nowak, P. M., Panigassi, A. P., Lima, A., Araujo Júnior, E., Nardozza, L., & Pares, D. (2019). Validation of Fetal Medicine Foundation algorithm for prediction of pre-eclampsia in the first trimester in an unselected Brazilian population. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*, *32*(2), 286–292. https://doi.org/10.1080/14767058.2017.1378332.
- Mallampati, D., Grobman, W., Rouse, D. J., & Werner, E. F. (2019). Strategies for Prescribing Aspirin to Prevent Preeclampsia: A Cost-Effectiveness Analysis. *Obstetrics and gynecology*, *134*(3), 537–544. https://doi.org/10.1097/AOG.0000000000003413
- Mallampati, D., & Werner, E. F. (2020). Response to Letter. *Obstetrics and gynecology*, *135*(1), 217–218. https://doi.org/10.1097/AOG.000000000003632
- Mone, F., O'Mahony, J. F., Tyrrell, E., Mulcahy, C., McParland, P., Breathnach, F., ··· McAuliffe, F. M. (2018). Preeclampsia Prevention Using Routine *Versus* Screening Test-Indicated Aspirin in Low-Risk Women. A Cost-Effectiveness analysis. *Hypertension*, 72(6), 1391–1396. https://doi.org/10.1161/HYPERTENSIONAHA.118.11718.
- Mosimann, B., Pfiffner, C., Amylidi-Mohr, S., Risch, L., Surbek, D., & Raio, L. (2017). First trimester combined screening for preeclampsia and small for gestational age a single centre experience and validation of the FMF screening algorithm. *Swiss medical weekly*, *147*, w14498. https://doi.org/ 10.4414/smw.2017.14498.
- NICE (2019). *Hypertension in pregnancy: diagnosis and management.* NICE guideline [NG133]. https://www.nice.org.uk/guidance/ng133.
- O'Gorman, N., Wright, D., Poon, L. C., Rolnik, D. L., Syngelaki, A., Wright, A., ··· Nicolaides, K. H. (2017b). Accuracy of competing-risks model in screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 49*(6), 751–755. https://doi.org/10.1002/uog.17399.
- O'Gorman, N., Wright, D., Poon, L. C., Rolnik, D. L., Syngelaki, A., de Alvarado, M., ··· Nicolaides, K. H. (2017a). Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation: comparison with NICE guidelines and ACOG recommendations. *Ultrasound in obstetrics & gynecology:* the official journal of the International Society of Ultrasound in Obstetrics and

- Gynecology, 49(6), 756-760. https://doi.org/10.1002/uog.17455.
- Oliveira, N., Magder, L. S., Blitzer, M. G., & Baschat, A. A. (2014). First-trimester prediction of pre-eclampsia: external validity of algorithms in a prospectively enrolled cohort. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 44*(3), 279–285. https://doi.org/10.1002/uog.13435
- Organización Mundial de la Salud. (2014). Recomendaciones de la OMS para la prevención y el tratamiento de la preeclampsia y la eclampsia. Ginebra: Organización Mundial de la Salud. https://apps.who.int/iris/bitstream/handle/10665/138405/9789243548333_spa.pdf;jsessionid=2A44A60E 4D64C613071BD6C684DFA5BB?sequence=1
- Ortved, D., Hawkins, T. L., Johnson, J. A., Hyett, J., & Metcalfe, A. (2019). Cost-effectiveness of first-trimester screening with early preventative use of aspirin in women at high risk of early-onset pre-eclampsia. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 53*(2), 239–244. https://doi.org/10.1002/uog.19076.
- Park, F., Deeming, S., Bennett, N., & Hyett, J. (2021). Cost-effectiveness analysis of a model of first-trimester prediction and prevention of preterm pre-eclampsia compared with usual care. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 58*(5), 688–697. https://doi.org/ 10.1002/uog.22193.
- Park, F. J., Leung, C. H., Poon, L. C., Williams, P. F., Rothwell, S. J., & Hyett, J. A. (2013). Clinical evaluation of a first trimester algorithm predicting the risk of hypertensive disease of pregnancy. *The Australian & New Zealand journal of obstetrics & gynaecology*, *53*(6), 532–539. https://doi.org/10.1111/ajo.12126.
- Poon, L. C., Syngelaki, A., Akolekar, R., Lai, J., & Nicolaides, K. H. (2013). Combined screening for preeclampsia and small for gestational age at 11-13 weeks. *Fetal diagnosis and therapy*, *33*(1), 16–27. https://doi.org/10.1159/000341712.
- Prasad, S., Sahota, D. S., Vanamail, P., Sharma, A., Arora, S., & Kaul, A. (2021). Performance of Fetal Medicine Foundation algorithm for first trimester preeclampsia screening in an indigenous south Asian population. *BMC pregnancy and childbirth*, *21*(1), 805. https://doi.org/ 10.1186/s12884-021-04283-6.
- Rezende, K., Cunha, A., Amim Junior, J., & Bornia, R. G. (2019a). External validation of the Fetal Medicine Foundation algorithm for the prediction of preeclampsia in a Brazilian population. *Pregnancy hypertension*, *17*, 64–68. https://doi.org/10.1016/j.preghy.2019.05.006
- Rezende, K., Cunha, A., Amim, J., Jr, Oliveira, W. M., Leão, M., Menezes, M., ··· Bornia, R. G. (2019). Performance of Fetal Medicine Foundation Software for Pre-Eclampsia Prediction Upon Marker Customization: Cross-Sectional Study. *Journal of medical Internet research*, *21*(11), e14738. https://doi.org/

- 10.2196/14738.
- Roberge, S., Nicolaides, K., Demers, S., Hvett, J., Chaillet, N., & Buiold, E. (2017). The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. *American iournal of obstetrics and gynecology*, *216*(2), 110–120.e6. https://doi.org/10.1016/j.ajog.2016.09.076.
- Roberge, S., Villa, P., Nicolaides, K., Giguère, Y., Vainio, M., Bakthi, A., ··· Buiold, E. (2012). Early administration of low-dose aspirin for the prevention of preterm and term preeclampsia: a systematic review and meta-analysis. *Fetal diagnosis and therapy*, *31*(3), 141–146. https://doi.org/10.1159/000336662
- Rocha. R. S., Gurgel Alves. J. A., Bezerra Maia e Holanda Moura. S., Arauio Júnior. E., Martins, W. P., Vasconcelos, C., ... Oriá, M. (2017). Comparison of three algorithms for prediction preeclampsia in the first trimester of pregnancy. *Pregnancy hypertension*, 10, 113–117. https://doi.org/10.1016/j.preghy.2017.07.146
- Rolnik, D. L., Wright, D., Poon, L., Syngelaki, A., O'Gorman, N., de Paco Matallana, C., ··· Nicolaides. K. H. (2017). ASPRE trial: performance of screening for preterm preeclampsia. *Ultrasound in obstetrics & gynecology: the official iournal of the International Society of Ultrasound in Obstetrics and Gynecology*, *50*(4), 492–495. https://doi.org/10.1002/uog.18816
- Rolnik, D. L., Wright, D., Poon, L. C., O'Gorman, N., Svngelaki, A., de Paco Matallana, C., ... Nicolaides, K. H. (2017b). Aspirin *versus* Placebo in Pregnancies at High Risk for Preterm Preeclampsia. *The New England journal of medicine*, *377*(7), 613–622. https://doi.org/10.1056/NEJMoa1704559
- Scazzocchio, E., Figueras, F., Crispi, F., Meler, E., Masoller, N., Mula, R., & Gratacos, E. (2013). Performance of a first-trimester screening of preeclampsia in a routine care low-risk setting. *American iournal of obstetrics and gynecology*, *208*(3), 203.e1–203.e10. https://doi.org/10.1016/j.ajog.2012.12.016
- Skråstad, R. B., Hov, G. G., Blaas, H. G., Romundstad, P. R., & Salvesen, K. Å. (2015). Risk assessment for preeclampsia in nulliparous women at 11-13 weeks gestational age: prospective evaluation of two algorithms. *BJOG: an international iournal of obstetrics and gynaecology*, *122*(13), 1781–1788. https://doi.org/ 10.1111/1471-0528.13194.
- Tan, M. Y., Wright, D., Svngelaki, A., Akolekar, R., Cicero, S., Janga, D., ··· Nicolaides, K. H. (2018). Comparison of diagnostic accuracy of early screening for preeclampsia by NICE guidelines and a method combining maternal factors and biomarkers: results of SPREE. *Ultrasound in Obstetrics & Gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 51(6), 743–750. https://doi.org/10.1002/uog.19039.
- Townsend, R., Khalil, A., Premakumar, Y., Allotev, J., Snell, K., Chan, C., Chappell, L. C., Hooper, R., Green, M., Mol, B. W., Thilaganathan, B., Thangaratinam, S., & IPPIC Network (2019). Prediction of pre-eclampsia: review of reviews. *Ultrasound in*

- obstetrics & gvnecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 54(1), 16–27. https://doi.org/10.1002/uog.20117.
- Urrutia, G. & Bonfill, X. (2010). Declaración PRISMA: una propuesta para meiorar la publicación de revisiones sistemáticas v metanálisis. *Medicina Clínica*, *135*, 507-511. https://doi.org/ 10.1016/j.medcli.2010.01.015.
- US Preventive Services Task Force (USPSTF), Bibbins-Domingo, K., Grossman, D. C., Currv, S. J., Barrv, M. J., Davidson, K. W., ··· Tseng, C. W. (2017). Screening for Preeclampsia: US Preventive Services Task Force Recommendation Statement. *JAMA*, *317*(16), 1661–1667. https://doi.org/10.1001/jama.2017.3439.
- US Preventive Services Task Force (USPSTF). Davidson KW. Barry MJ. Mangione. C.M.. Cabana. M.. Caughev. A.B. *et al.* (2021). Aspirin Use to Prevent Preeclampsia and Related Morbidity and Mortality: US Preventive Services Task Force Recommendation Statement. *JAMA*, *326* (12), 1186-1191. https://doi.org/10.1001/jama.2021.14781
- Vargas H, Víctor Manuel, Acosta A, Gustavo, & Moreno E, Mario Adán. (2012). La preeclampsia un problema de salud pública mundial. *Revista chilena de obstetricia y ginecología*, 77(6), 471-476. https://doi.org/10.4067/S0717-75262012000600013
- Werner, E. F., Hauspurg, A. K., & Rouse, D. J. (2015). A Cost-Benefit Analysis of Low-Dose Aspirin Prophylaxis for the Prevention of Preeclampsia in the United States. *Obstetrics and gynecology*, *126*(6), 1242–1250. https://doi.org/10.1097/AOG.00000000001115
- Wright, D., & Nicolaides, K. (2021). Re: Implementation of routine first trimester combined screening for pre-eclampsia: a clinical effectiveness study. *BJOG: an international journal of obstetrics and gynaecology*, *128*(1), 141–142. https://doi.org/10.1111/1471-0528.16445
- Zamora Romero, J., Plana, M.N., Abraira Santos, V. (2009). Estudios de evaluación de la validez de una prueba diagnóstica: revisión sistemática y metaanálisis. Nefrología 29, 15-20. https://doi.org/10.3265/NEFROLOGIA.2009.29.S.E.noID.36.free.
- Zhang, M., Wan, P., Ng, K., Singh, K., Cheng, T. H., Velickovic, I., ··· Wlody, D. (2020). Preeclampsia among African American Pregnant Women: An Update on Prevalence, Complications, Etiology, and Biomarkers. *Obstetrical & gynecological survey*, *75*(2), 111–120. https://doi.org/10.1097/OGX.0000000000000747.

Anexos

Anexo 1: estrategias de búsqueda

Ovid MEDLINE(R) <1946 to July Week 5 2021>

- 1. *Pre-Eclampsia/di, ep, pc or *Pregnancy Trimester, First/
- 2. preeclampsia.m_titl.
- 3. "pre-eclampsia".m_titl.
- 4. 2 or 3
- 5. 1 or 4
- 6. ("fmf" adj2 (algorithm or software)).ti. or "FMF2012".ti,ab.
- 7. "fetal medicine foundation".ti,ab.
- 8. 6 or 7
- 9. algorithms/ (271328)
- 10. Software/ or Software Validation/
- 11. 9 or 10
- 12. 8 or 11
- 13. 5 and 12
- 14. "sensitivity and specificity"/ or "predictive value of tests"/ or roc curve/
- 15. "reproducibility of results"/
- 16. area under curve/
- 17. (diagnos* or sensitivity or specificity or (predictive adj3 value*) or "roc curve" or "receiver operating curve" or "AUC" or "area under curve" or "reproducibility of results" or concordance or "false positive" or "false negative").ti,ab.
- 18. 14 or 15 or 16 or 17
- 19. 13 and 18

Embase

.....

- 1. 'preeclampsia'/mj/exp/dm_di,dm_ep,dm_pc
- 2. 'first trimester pregnancy'/exp/mj
- 3. 'pre eclampsia':ti OR preeclampsia:ti
- 4. 1 OR 2 OR 3
- 5. 'fetal medicine foundation':ti,ab
- 6. ((fmf NEXT/2 (algorithm OR software)):ti) OR fmf2012:ti,ab
- 7. ('algorithm'/exp OR 'software'/de OR 'software validation'/de) AND (fmf:ti,ab OR fmf2012:ti,ab)
- 8. 5 OR 6 OR 7
- 9. 4 AND 8
- 10. 9 AND [embase]/lim
- 11. diagnos*:ti,ab OR sensitivity:ti,ab OR specificity:ti,ab OR ((predictive NEXT/3 value*):ti,ab) OR 'roc curve':ti,ab OR 'receiver operating curve':ti,ab OR 'auc':ti,ab OR 'area under curve':ti,ab OR 'reproducibility of results':ti,ab OR concordance:ti,ab OR 'false positive':ti,ab OR 'false negative':ti,ab

- 12. 'sensitivity and specificity'/de OR 'predictive value'/de OR 'receiver operating characteristic'/de OR 'reproducibility'/de OR 'area under the curve'/exp
- 13. 11 OR 12
- 14. 10 AND 13

Cochrane Library
Preeclampsia and pregnancy and risk assessment and screening

Anexo 2: estudios excluidos tras la lectura a texto completo

Artículos excluido	os tras lectura a texto completo
Artículo	Motivo de exclusión
Poon et al., 2013	Intervención (combinación de dos algoritmos)
Rocha <i>et al.</i> , 2017	Intervención (los algoritmos estudiados no utilizan los mismos parámetros que el de la FMF). El objetivo es validar un nuevo algoritmo basado en el de FMF pero más sencillo.
Lakovschek et al., 2018	Resultados. No aportan datos de discriminación diagnóstica. No comparan los resultados del test con los del estándar (resultado del embarazo).
Rezende <i>et al.</i> , 2019a	Intervención (utilizan el algoritmo de la FMF pero la combinación de factores de riesgo no coincide con ninguno de los modelos incluidos en el análisis de este informe. Utiliza la CM junto a MAP+UtAPI (factores biofísicos).
Rezende <i>et al.</i> , 2019b	No es un estudio de cribado propiamente. Este estudio en concreto valora si el factor materno de la etnia modifica o no los resultados del algoritmo e la FMF2012.
Bruneli <i>et al.,</i> 2020	Intervención (los algoritmos estudiados no utilizan los mismos parámetros que el de la FMF). El objetivo es validar un nuevo algoritmo basado en el de FMF pero más sencillo.
He et al., 2020	Intervención (un algoritmo diferente al de la FMF y que utiliza otras variables: muesca diastólica, presión arterial uterina y la PAPP-A).
Cordisco et al., 2021	Tipo de resultado. Estudio de efectividad clínica del cribado que aporta datos no ajustados por el tratamiento preventivo (aspirina). No es posible valorar el rendimiento real de la prueba.
Guy et al., 2021	Tipo de resultado. Estudio de efectividad clínica del cribado que aporta datos no ajustados por el tratamiento preventivo (aspirina). No es posible valorar el rendimiento real de la prueba

Anexo 3: Características de los estudios incluidos en la revisión sistemática

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
Park 2013	Australia	2016-2019	Embarazos de un feto en 1er trimestre de gestación de un cribado de alteraciones cromosómicas. No se excluyeron tratamientos antihipertensivos, aspirina o cualquier otro que pueda afectar la prevalencia de la preeclampsia.	Embarazos múltiples	3.099	3.014 (97,26%)	ISSHP 2001	Poon 2010 Características maternas + MAP+UtAPI+ PAPP-A	10% FP
Oliveira 2014	Estados Unidos	2007-2010	Mujeres con gestación única a las 9-14 sema- nas de gestación con consentimiento informado firmado.	Embarazos termi- nados antes de la 20 semanas de gestación	2.969	2.833 (95,42%)	ACOG 2002	Poon 2010 Características maternas + MAP+UtAPI+ PAPP-A	10% FP
Skråstad 2014	Noruega	2010-2012	Mujeres nulíparas con 11-13 semanas de gestación con consenti- miento informado firmado.	Tratamiento con medicamentos anticoagulantes o aspirina durante el embarazo.	560	541 (96,6%)	ACOG 2002 modificada. (proteinuria ≥0,3g/24h medida más de una vez en intervalos de 4 a 6 horas)	First Trimester Screening Program versión 2.8 Características maternas + MAP+UtAPI+PIGF+ PAPP-A	10% FP

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
Mosimann 2017	Suiza	2012-2016	Embarazos de un feto que aceptaron el cribado de PE	No descritos	1.372	1.129 (82,28%)	ISSHP 2014	Algoritmo de la FMF Características maternas + MAP+UtAPI+ PIGF	Riesgo PE temprana: ≥200 10% FP
Rolnik 2017	Reino Unido España Italia Bélgica Grecia Israel	abril-junio 2014 julio 2015- abril 2016	Embarazos de un feto con 11-13 semanas de gestación. Edad materna ≥ 18 años, sin enfermedad mental grave o dificultad de aprendizaje y embarazo único con feto vivo sin anomalías importantes demostradas en la exploración de 11 a 13 semanas	Embarazos sin seguimiento y los que terminaron en interrupción o aborto espontáneo.	26.941	25.797 (95,75%)	ISSHP 2001	Características maternas + MAP+UtAPI+PIGF+ PAPP-A	Riesgo PE pretérmino: ≥100
O´Gorman 2017	Reino Unido España Italia Bélgica Grecia	Febrero- septiembre 2015	Embarazos de un feto con 11-13 semanas de gestación	nd	nd	8.775 (-)	ISSHP 2001	Características maternas + MAP+UtAPI+PIGF+ PAPP-A	10% FP
Guizani 2018	Bélgica	2011-2013	Mujeres con embarazos únicos que entraron en el centro para las pruebas de detección del síndrome de Down, ≥ 18 años, sin discapacidad mental	embarazos múltiples, con aneuploidías o anomalías fetales importantes, o que terminaron en interrupción o aborto espontáneo.	3.935	3.239 (82,31%)	ISSHP 2001	Características maternas + MAP+UtAPI+PIGF+ PAPP-A	10% FP

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
			grave, embarazo único con un feto vivo en la exploración de 11 a 13 semanas y parto posterior de un bebé vivo fenotípicamente normal o muerte fetal ≥ 24 semanas de gestación.	Antecedentes de EP que ya estaban tomando aspirina antes de asistir a la exploración de 11 a 13 semanas.					
Tan 2018	Reino Unido (Inglaterra)	Abril- diciembre 2016	Mujeres ≥ 18 años de edad, con embarazo único y feto vivo en la exploración de 11 a 13 semanas.	Mujeres inconscientes o gravemente enfermas, con dificultades de aprendizaje o enfermedades mentales graves. Anomalías fetales importantes identificadas en la exploración de 11 a 13 semanas.	17.051	16 747 (98,22%)	ISSHP 2001 ACOG 2013	Mini test: Características maternas + MAP +PAPP-A Triple test: Características maternas + MAP+UtAPI+ PIGF	10 % tasa de cribado positivo
Di Martino 2019	Italia	2014-2017	Embarazo único, edad materna > 18 años, feto vivo a las 11-13 semanas de gestación y consentimiento informado por escrito. Y con seguimiento completo del embarazo.	Falta de seguimiento registrado en la base de datos electrónica, gestación múltiple, anomalías congénitas fetales y aborto espontáneo antes de las 24 semanas de gestación.	12.284	11.632 (94,69%)	ISSHP 2001	Características maternas	10% FP

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
Lobo 2019	Brasil		Embarazo único con feto de tamaño (CRL) de 45-84 mm	(1) incapacidad para realizar un estudio Doppler abdominal de las arterias uterinas, (2) fetos con anomalías cromosómicas o defectos estructurales graves (3) uso regular de ácido acetilsalicílico durante el embarazo, (4) gestación que terminó en aborto espontáneo o muerte fetal antes de la semana 24, (5) pérdidas durante el seguimiento y (6) incapacidad para clasificar el trastorno hipertensivo	820	617 (75,24%)	Presión arterial sistólica > 140 mm Hg y / o presión arterial diastólica > 90 mm Hg en al menos dos ocasiones con 6 h de diferencia en un paciente previamente normotenso asociado a proteinuria de 300 mg/24 h	First Trimester Screening Program versión 2.8.0 Características maternas + MAP+UtAPI+ PIGF+PAPP-A	10% FP
Chaemsaithong 2019	Japón China Taiwan India Tailandia Singapur	2016-2018	Embarazos únicos con feto vivo de 11-13 +6 semanas de gestación. Edad materna ≥18 años.	Mujeres inconscientes o gravemente enfer- mas, con dificultades de aprendizaje o enfermedades menta- les graves. Anomalías fetales importantes identificadas en la	nd	10 935 (-)	ISSHP 2001 ACOG 2013	FMF triple test: Características maternas + MAP+UtAPI+ PIGF	10%FP

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
				exploración de 11 a 13 semanas	(117)	(///			
Boutin 2021	Canadá	2011-2014	Mujeres nulíparas con embarazos únicos en 11+0 a 13+6 semanas y seguimiento completo	Embarazos múltiples, anomalías letales fetales y anomalías cromosómicas fetales que condujeran a la interrupción médica del embarazo y la muerte fetal antes de las 14 semanas de gestación. Mujeres que informaron haber tomado aspirina diariamente durante el transcurso de sus embarazos (cualquier indicación).	4.738	4.575 (96,6%)	Canadian Hypertensive disorders of Pregnancy Group 2014	Características maternas + MAP+UtAPI+ PIGF+PAPP-A	Test positivo riesgo ≥1:70 10%FP
Goto 2021	Japón	2017-2019	Embarazos únicos con feto vivo y sin anomalías de 11-13 semanas de gestación. Edad materna ≥18 años.	Mujeres con dificultades de aprendizaje o enfermedades mentales graves. Abortos espontáneos o interrupciones del embarazo antes de la semana 22 de gestación.	1.036	913 (88,13%)	Japan Society for the Study of Hypertension in Pregnancy 2005	FMF triple test: Características maternas + MAP+UtAPI+PIGF	10% FP
Chen 2021	Taiwan	2017-2018	Mujeres ≥ 18 años con	Embarazos con	nd	700 (-)	ACOG 2013	FMF triple test:	10% FP

Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
			embarazos únicos en	factores maternos,				Características	
			11+0 a 13+6 semanas de	como enfermedad				maternas +	
			gestación y que dieron a	mental y grave.				MAP+UtAPI+ PIGF	
			luz a un nacido vivo o	Anomalías fetales					
			hubo muerte fetal a	diagnosticadas en el					
			partir de las 24 semanas	momento de la					
			de gestación.	selección.					
lu 2021	China	2017-2019	Mujeres ≥ 18 años con	Mujeres que no	11 741	10 899	ISSPH 2018	FMF triple test:	10% FP
			embarazo único, de	pudiera dar su		(92,83%)	ACOG 2013	Características	
			etnia china materna y	consentimiento				maternas +	
			feto vivo a las 11 + 0 a 13	informado completo				MAP+UtAPI+PIGF	
			+ 6 semanas de	por cualquier motivo,					
			gestación.	como una enfermedad				Características	
				grave, dificultades de				maternas +	
				aprendizaje o una				MAP+UtAPI+PIGF+	
				enfermedad mental				PAPP-A	
				importante. Además,					
				se excluyeron los					
				embarazos con					
				anomalías					
				cromosómicas o					
				estructurales					
				importantes del feto y					
				los que se					
				interrumpieron, así					
				como los que					
				terminaron en aborto					
				espontáneo o muerte					
				fetal <24 semanas					
				completas de					

Características	Características de los estudios incluidos en la revisión sistemática								
Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte
				gestación					
Prasad 2021	India	2017-2019	Mujeres con embarazos únicos que asistían para su evaluación del riesgo de síndrome de Down en el primer trimestre	Anomalías fetales importantes en el momento de la exploración o un aborto sin diagnóstico.	1.975	1.863 (94,33%)	ISSPH 2001	Características maternas + MAP+ UtAPI+PIGF+ PAPP-A	10% FP

Nd: no descrito FP: falsos positivos

*ISSHP 2001: Hipertensión de novo que surge después de las 20 semanas de gestación, que vuelve a la normalidad después del parto, con proteinuria (orina de 24 h ≥300 mg o relación proteína/creatinina en orina puntual ≥30 mg proteína/mmol de creatinina)

ISSHP 2014/ISSHP 2018: Hipertensión que se desarrolla después de las 20 semanas de gestación y la coexistencia de una o más de las siguientes condiciones de nueva aparición: 1. Proteinuria 2. Otra disfunción orgánica materna: insuficiencia renal (creatinina >90 umol/L), compromiso hepático (transaminasas elevadas y/o dolor abdominal de intensidad severa, en epigástrico o en el cuadrante superior derecho), complicaciones neurológicas (los ejemplos incluyen eclampsia, estado mental alterado, ceguera, accidente cerebrovascular, hiperreflexia con clonus, cefalea intensa con hiperreflexia, escotomas persistentes), complicaciones hematológicas (trombocitopenia, coagulación intravascular diseminada(CID), hemólisis) 3. Disfunción uteroplacentaria (restricción del crecimiento fetal)

ACOG 2002: • Presión arterial sistólica ≥ 140 mm Hg o diastólica ≥ 90 mm Hg que ocurre después de 20 semanas de gestación en una mujer con presión arterial previamente normal • Proteinuria, definida como excreción urinaria ≥ 0,3 g proteína en una muestra de orina de 24 horas.

ACOG 2013 Presión arterial sistólica de ≥140 mmHg o diastólica de ≥90 mmHg en al menos dos ocasiones con 4 h de diferencia, desarrollada después de 20 semanas de gestación en mujeres normotensas más: proteinuria (≥300 mg/24 h o relación proteína/creatinina ≥30 mg/mmol o ≥2+ en la prueba con tira reactiva). En ausencia de proteinuria hipertensión de nueva aparición con al menos una de las siguientes condiciones: insuficiencia renal (creatinina sérica >1,1 mg/dl o aumento del doble de la creatinina sérica en ausencia de una enfermedad renal subyacente), afectación hepática (concentración sanguínea de transaminasas al doble del nivel normal), complicaciones neurológicas.

síntomas visuales), trombocitopenia (recuento de plaquetas <100 000/μl) o edema pulmonar.

NICE 2019: Hipertensión de nueva aparición (sistólica superior a 140 mmHg o diastólica superior a 90 mmHg) después de las 20 semanas de embarazo y la coexistencia de 1 o más de las siguientes condiciones de nueva aparición:1 proteinuria (relación proteína:creatinina en orina de 30 mg/mmol o más o albúmina:creatinina de 8 mg/mmol o más, o al menos 1 g/litro [2+] en la prueba con tira reactiva). 2. Otra disfunción orgánica materna: insuficiencia renal (creatinina 90 micromol/litro o más, 1,02 mg/100 ml o más), compromiso hepático (transaminasas elevadas [alanina aminotransferasa o aspartato aminotransferasa por encima de 40 Ul/litro] con o sin dolor abdominal epigástrico o en el cuadrante superior derecho), Complicaciones neurológicas como eclampsia, alteración del estado mental, ceguera, ictus, clonus, cefaleas intensas o escotomas visuales persistentes., complicaciones hematológicas como trombocitopenia (recuento de plaquetas por debajo de 150.000/microlitro), coagulación intravascular diseminada o hemólisis. 3. disfunción uteroplacentaria, como restricción del crecimiento fetal, análisis anormal de la forma de onda doppler de la arteria umbilical o muerte fetal.

Canadian Hypertensive disorders of Pregnancy Group 2014: hipertensión observada después de las 20 semanas de gestación en presencia de proteinuria o una condición adversa o una

Características de los estudios incluidos en la revisión sistemática									
Estudio	País	Fecha cribado	Criterios inclusión	Criterios exclusión	Población cribado (N)	Población analizada N (%)	Definición de PE*	Algoritmo usado	Punto de corte

complicación grave (como cefalea, dolor torácico, disnea, saturación de oxígeno <97%, recuento elevado de glóbulos blancos, índice internacional normalizado (INR) elevado o parcial activado). tiempo de tromboplastina o recuento bajo de plaquetas, creatinina sérica alta o ácido úrico, náuseas y dolor epigástrico).

Japan Society for the Study of Hypertension in Pregnancy: Hipertensión de novo acompañada de proteinuria u otras disfunciones orgánicas (, la EP se diagnostica por la presencia de hipertensión de novo después de las 20 semanas de gestación acompañada de proteinuria y/o evidencia de lesión renal aguda materna, disfunción hepática, características neurológicas, hemólisis o trombocitopenia, o restricción del crecimiento fetal.

Anexo 4: Descripción de las poblaciones de los estudios incluidos en la revisión sistemática

Características maternas incluidas en la calculadora de la FMF	Park 2013	Oliveira 2014	Skråstad 2014	Mosimann 2017	O 'Gorman 2017	Rolnik 2017	Guizani 2017	Tan 2018	Lobo 2019	Di Martino 2019	Chaemsaithong 202019	Goto 2021	Boutin 2021	Chen 2021	Hu 2021	Prasad 2021
Edad en años mediana(RIQ)	31-34 [*]	27-30*	25-27	31 (27-35)	30-31*	31,7 (27,7-35,2)	30-32*	31,5 (27,4-35,1)	30 (25-35)	32-34*	32-35	34-40	28,7 (26,1-31,3)	nd	29-30	30-31
IMC kg/m² mediana (RIQ)	23-26*	26-31*	24-25	22,8 (20,6-26,2)	24-28*	24,7 (21,8-28,2)	25-26	24,7 (22-28,7)	26-29*	23-25	21-24	20-22	23,8 (21,6-26,9)	nd	22-23	24-26
Origen étnico (%) Blanco Negro Sur asiático Este asiático Mixto	67,6 32,1 10,1 20,6 0,6	43,0 50,0 4,5 0,5 0,2	98,0 0 0,2 0,5 0,9	73,5 14,3 5,2 4,4 2,6	78,5 12,5 1,7 5,3 2,1	79,0 12,1 4,6 2,0 2,3	82,7 2,9 1,54 0,25 0,2	72,3 14,4 8,3 2,5 2,6	- 13,9 - - 54,4	97,0 14,4 8,3 2,5 2,6	0,19 0 2,78 97 0	0 0 0 100 0	94,9 1,3 0,8 3,1 0	0 0 0 100 0	0 0 0 100 0	0 0 100 0
Fumador (%)	3,0	9,5	11,3	8,6	8,3	8,0	8,3	6,8	15,4	nd	4,3	5,6	nd	nd	0,04	nd
Antecedentes familiares PE (%)	7,3	4,9	6,5	1,2	5,2	3,3	1,3	3,2	8,2	2,6	1,5	3,8	nd	nd	0,13	nd
Método de concepción: (%) Espontáneo Estimulación ovárica Fertilización <i>in vitro</i>	95 2 3	nd 1 nd	91 4,4 4,6	92 - 8	97 0,7 2,6	96,4 0,6 3,0	96,7 1,6 1,6	95,8 0,8 3,4	nd	96,3 - 3,7	92,6 1,5 5,9	- - 23,3	93,7 4,5 1,8	nd	98,1 - 1,9	- - 5,2
Hipertensión crónica (%)	1,2	8,4	0,2	1,9	1,1	1,2	1,5	0,85	10,5	0,83	0,53	0,87	0,3	nd	0,84	1,6
Diabetes I (%) Diabetes II (%)	nd nd	4,04	0	0,7	0,8	0,8	0,59	0,71	3,7	3,17	0,26	0,44	0,02 0,10	nd nd	0,31	2,8
Lupus eritematoso sistémico (%)	nd	nd	0	0,5	0,4	0,5	0,15	0,24	2,6	0,42	0,38	nd	nd	nd	0,21	0,3
Síndrome antifosfolipídico (%)	nd	nd	0	0,5	0,4	0,5	0,15	0,24	2,0	U, 4 Z	0,36	nd	0,1	nd	0,21	0,3
Nulíparas (%)	53	43	100	51	47	47,2	35,5	46,1	41,3	65,7	57,7	53,4	100	nd	69,6	60,2

Anexo 5: Análisis del efecto umbral en la estimación del riesgo "a priori". Modelo 1

Coeficiente de correlación de Spearman: -0,571 p = 0,139

(Logit(TPR) vs Logit(FPR)

Modelo de Moses (D = a + bS)

Regresión ponderada (Inversa de la Varianza)

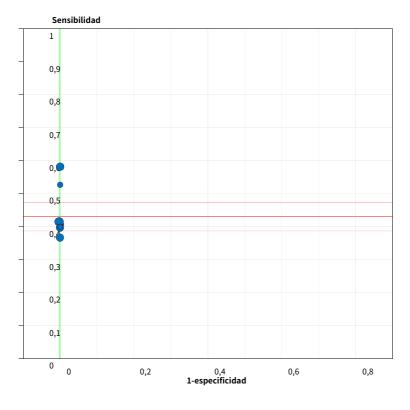
Var	Coeff.	Std. Error	Т	p-value
a	4,374	0,095	45,935	0,0000
b(1)	0,985	0,038	25,915	0,0000

Nº de estudios = 8

Filtro OFF

Se añade 1/2 solo a los estudios con casillas con ceros

PLANO ROC



Sensibilidad

0,43 (IC 95% 0,39-0,47) Chi-cuadrado = 10,19; gl = 7 (ρ = 0,1778)

Especificidad

0,90 (IC95% 0,90-0,90) Chi-cuadrado = 0,99; gl = 7 (ρ = 0,9949)

Anexo 6: Análisis del efecto umbral en la estimación del riesgo Modelo 2: Mini test

Coeficiente de correlación de Spearman: 0,000 p = 1,000

(Logit(TPR) vs Logit(FPR)

Modelo de Moses (D = a + bS)

Regresión ponderada (Inversa de la Varianza)

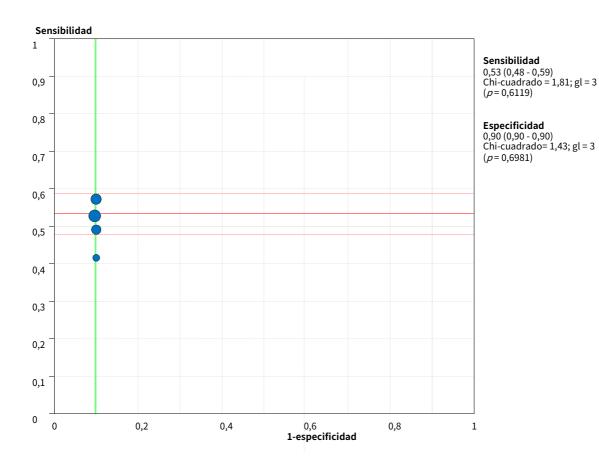
Var	Coeff.	Std. Error	Т	p-value
a	4,295	0,383	11,201	0,0079
b(1)	0,935	0,184	5,091	0,0365

 N^0 de estudios = 12

Filtro OFF

Se añade 1/2 solo a los estudios con casillas con ceros

PLANO ROC



Anexo 7: Análisis del efecto umbral en la estimación del riesgo del Triple test. Modelo 3.

Coeficiente de correlación de Spearman: 0,085 p = 0,781

(Logit(TPR) vs Logit(FPR)

Modelo de Moses (D = a + bS)

Regresión ponderada (Inversa de la Varianza)

Var	Coeff.	Std. Error	Т	p-value
a	4,285	0,179	23,927	0,0000
b(1)	0,904	0,122	7,390	0,0000

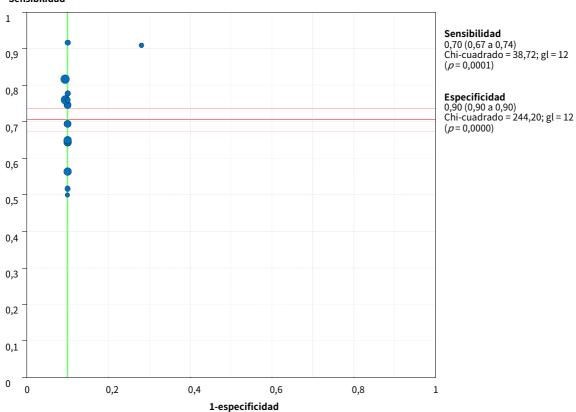
 N^0 de estudios = 13

Filtro OFF

Se añade 1/2 solo a los estudios con casillas con ceros

PLANO ROC

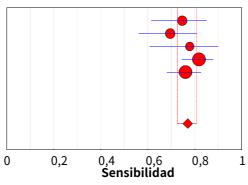




Anexo 8: Análisis de sensibilidad del Triple test. Modelo 3.

Por poblaciones

EUROPEA

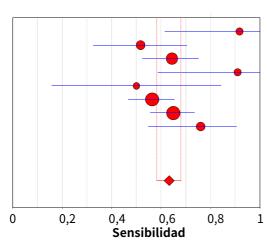


Sensibilidad (95% CI)

O'Gorman 2017	0,75	(0,62 - 0,85)
O'Gorman2017	0,69	(0,56 - 0,81)
Guizani 2018	0,78	(0,61 - 0,90)
Tan 2018	0,82	(0,74 - 0,88)
Tan 2018	0.76	(0.68 - 0.83)

Sensibilidad ponderada = 0,77 (0,73 - 0,81) Chi-cuadrado = 3,89; gl = 4 (p= 0,4212) Inconsistencia (l^2) = 0,0 %

NO EUROPEA

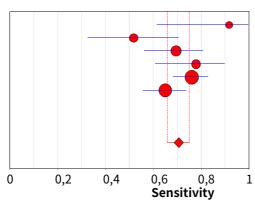


Sensibilidad (95% CI)

Park 2013	0,92	(0,62 - 1,00)
Oliveira 2014	0,52	(0,33 - 0,71)
Chaemsaithong 2019	0,64	(0,52 - 0,75)
Goto 2021		(0,59 - 1,00)
Chen 2021	0,50	(0,16 - 0,84)
Hu2021	0,56	
Hu 2021	0,65	(0,56 - 0,74)
Prasad 2021	0,76	(0,55 - 0,91)

Sensibilidad ponderada = 0,63 (0,58 - 0,68) Chi-cuadrado = 16,20; gl = 7 (p = 0,0234) Inconsistencia (l^2) = 56,8 %

Con PAPP-A

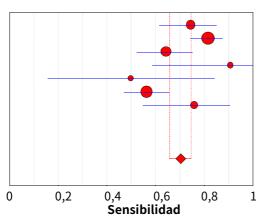


Sensibilidad (95% CI)

Park 2013 Oliveira 2014 O'Gorman Guizani 2018 Tan 2018	0,92 0,52 0,69 0,78 0,76	(0,62 - 1,00) (0,33 - 0,71) (0,56 - 0,81) (0,61 - 0,90) (0,68 - 0,83)
Tan 2018	0,76	(0,68 - 0,83)
Hu 2021	0,65	(0,56 - 0,74)

Sensibilidad ponderada = 0,71 (0,66 - 0,75) Chi-cuadrado = 12,61; gl = 5 (p = 0,0273) Inconsistencia (l^2) = 60,3 %

Con PIGF



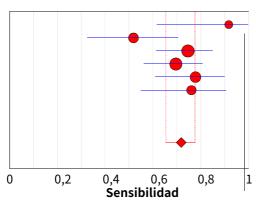
Sensibilidad (95% CI)

O'Gorman 2017		(0,62 - 0,85)
Tan 2018	0,82	(0,74 - 0,88)
Chaemsaithong 2019	0,64	(0,52 - 0,75)
Goto 2021	0,91	(0,59 - 1,00)
Chen 2021	0,50	(0,16 - 0,84)
Hu2021		(0,47 - 0,66)
Prasad 2021	0,76	(0,55 - 0,91)

Sensibilidad ponderada = 0,70 (0,66 - 0,75) Chi-cuadrado= 26,10; gl = 6 (p = 0,0002) Inconsistencia (l^2) = 77,0 %

Por definición de PE

Definición PE 2001

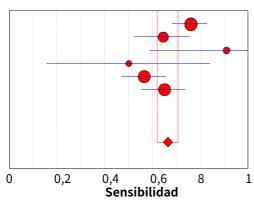


Sensibilidad (95% CI)

Park 2013	0,92	(0,62 - 1,00)
Oliveira 2014	0,52	(0,33 - 0,71)
O'Gorman 2017	0,75	(0,62 - 0,85)
O'Gorman2017	0,69	(0,56 - 0,81)
Guizani 2018	0,78	(0,61 - 0,90)
Prasad 2021	0,76	(0,55 - 0,91)

Sensibilidad ponderada = 0,72 (0,65 - 0,78) Chi-cuadrado = 9,42; gl = 5 (p= 0,0933) Inconsistencia (l^2) = 46,9 %

Definición 2014

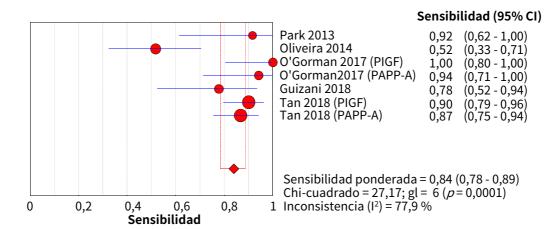


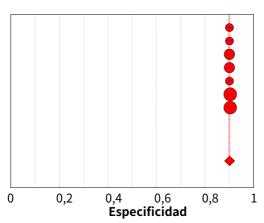
Sensibilidad (95% CI)

Tan 2018	0,76	(0,68 - 0,83)
Chaemsaithong 2019	0,64	(0,52 - 0,75)
Goto 2021	0,91	(0,59 - 1,00)
Chen 2021	0,50	(0,16 - 0,84)
Hu 2021	0,56	(0,47 - 0,66)
Hu 2021	0,65	(0,56 - 0,74)

Sensibilidad ponderada = 0,66 (0,62 - 0,71) Chi-cuadrado = 16,14; gl = 5 (ρ = 0,0065) Inconsistencia (l^2) = 69,0 %

Anexo 9: Rendimiento del triple test para la PE temprana

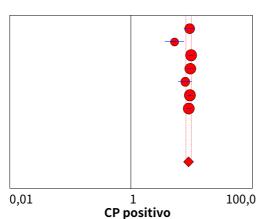




Especificidad (95% CI)

Park 2013	0,90	(0,89 - 0,91)
Oliveira 2014	0,90	(0,89 - 0,91)
O'Gorman 2017 (PIGF)	0,90	(0,89 - 0,91)
O'Gorman2017 (PAPP-A)	0,90	(0,89 - 0,91)
Guizani 2018	0,90	(0,89 - 0,91)
Tan 2018 (PIGF)	0,90	(0,90 - 0,91)
Tan 2018 (PAPP-A)	0,90	(0.90 - 0.91)

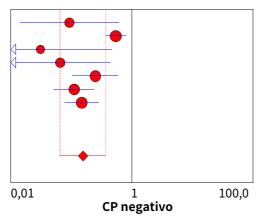
Especificidad ponderada = 0,90 (0,90 - 0,90) Chi-cuadrado = 1,30; gl = 6 (ρ = 0,9719) Inconsistencia (l^2) = 0,0 %



CP positivo (95% CI)

Park 2013	9.17	(7,50 - 11,22)
Oliveira 2014	5,18	(3,58 - 7,49)
O'Gorman 2017 (PIGF)	9,72	(8,79 - 10,74)
O'Gorman2017 (PAPP-A)	9,41	(8,23 - 10,76)
Guizani 2018	7,78	(5,95 - 10,17)
Tan 2018 (PIGF)	9,26	(8,42 - 10,20)
Tan 2018 (PAPP-A)	8,91	(7,99 - 9,94)

Modelo de efectos aleatorios CP positivo ponderado = 8,85 (7,99 - 9,80) Cochran-Q = 20,04; gl = 6 (p= 0,0027) Inconsistencia (l²) = 70,1 %



Negative LR (95% CI)

Park 2013	0,09	(0,01 - 0,60)
Oliveira 2014	0,54	(0,37 - 0,78)
O'Gorman 2017 (PIGF)	0,03	(0,00 - 0,47)
O'Gorman2017 (PAPP-A)	0,07	(0,01 - 0,44)
Guizani 2018	0,25	(0,10 - 0,59)
Tan 2018 (PIGF)	0,11	(0,05 - 0,24)
Tan 2018 (PAPP-A)	0.15	(0.08 - 0.28)

Modelo de efectos aleatorios CP negativo ponderado = 0,16 (0,06 - 0,37) Cochran-Q = 38,25; gl = 6 (p = 0,0000) Inconsistencia (l^2) = 84,3 %

Anexo 10: Análisis del efecto umbral en la estimación del riesgo del Cuádruple test. Modelo 4.

Coeficiente de correlación de Spearman: -0,399 p = 0,254

(Logit(TPR) vs Logit(FPR)

Modelo de Moses (D = a + bS)

Regresión ponderada (Inversa de la Varianza)

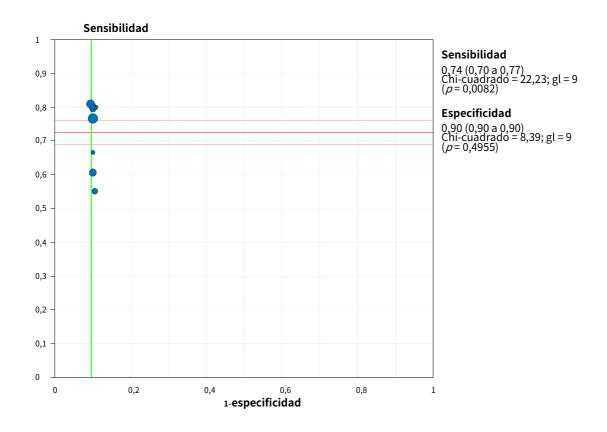
Var	Coeff.	Std. Error	Т	p-value
a	4,505	0,060	74,739	0,0000
b(1)	1,079	0,048	22,654	0,0000

Nº de estudios = 10

Filtro OFF

Se añade 1/2 solo a los estudios con casillas con ceros

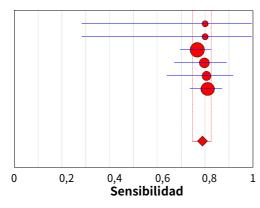
PLANO ROC



Anexo 11: Análisis de sensibilidad del Modelo4: Cuádruple test

Por poblaciones

EUROPEA

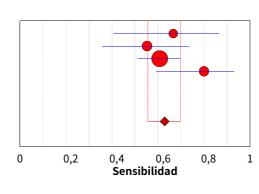


Sensibilidad (95% CI)

Skrastad 2014	0,80	(0,28 - 0,99)
Mosimann 2017	0,80	(0,28 - 0,99)
Rolnik 2017	0,77	(0,70 - 0,83)
O'Gorman 2017	0,80	(0,67 - 0,89)
Guizani 2018	0,81	(0,64 - 0,92)
Tan 2018	0,81	(0,74 - 0,87)

Sensibilidad ponderada = 0,79 (0,75 a 0,83) Chi-cuadrado = 1,00; gl = 5 (p = 0,9630) Inconsistencia (l^2) = 0,0 %

NO EUROPEA



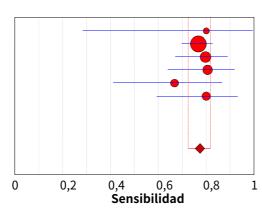
Sensibilidad (95% CI)

Lobo 2019	0,67	(0,41 - 0,87)
Boutin 2021	0,55	(0,36 - 0,74)
Hu 2021	0,61	(0,51 - 0,70)
Prasad 2021	0,80	(0,59 - 0,93)

Sensibilidad ponderada = 0,63 (0,56 a Chi-square = 4,52; gl = 3 (p = 0,2105) inconsistencia (l^2) = 33,6 %

Por definición de PE

Definición PE 2001

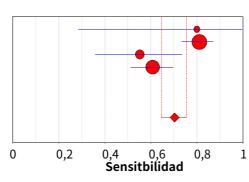


Sensibilidad (95% CI)

Skrastad 2014	0,80	(0,28 - 0,99)
Rolnik 2017	0,77	(0,70 - 0,83)
O'Gorman 2017	0,80	(0,67 - 0,89)
Guizani 2018	0,81	(0,64 - 0,92)
Lobo 2019	0,67	(0,41 - 0,87)
Prasad 2021	0,80	(0,59 - 0,93)

Sensibilidad ponderada = 0,77 (0,72 - 0,82) Chi-cuadrado = 1,64; gl = 5 (p= 0,8959) Inconsistencia (l^2) = 0,0 %

Definición PE 2014

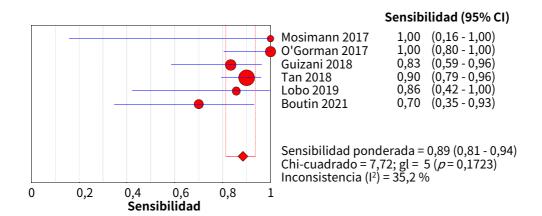


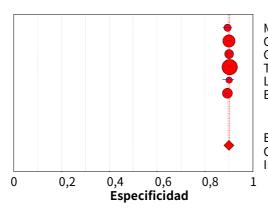
Sensibilidad (95% CI)

Mosimann 2017	0,80	(0,28 - 0,99)
Tan 2018		(0,74 - 0,87)
Boutin 2021	0,55	(0,36 - 0,74)
Hu 2021	0,61	(0,51 - 0,70)

Sensibilidad ponderada = 0,70 (0,65 - 0,75) Chi-cuadrado = 16,57; gl = 3 (ρ = 0,0009) Inconsistencia (I²) = 81,9 %

Anexo 12: Rendimiento del cuádruple test para la PE temprana

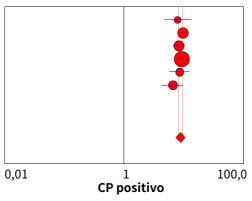




Especificidad (95% CI) 2017 0,89 (0,87 - 0,91)

MOSIMann 201 <i>1</i>	0,09	(0,01 - 0,91)
O'Gorman 2017	0,90	(0,89 - 0,91)
Guizani 2018	0,90	(0,89 - 0,91)
Tan 2018	0,90	(0,90 - 0,91)
Lobo 2019	0,90	(0,87 - 0,92)
Boutin 2021	0,89	(0,88 - 0,90)

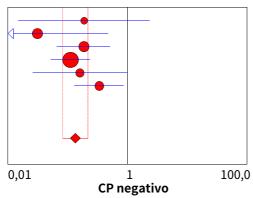
Especificidad ponderada = 0,90 (0,90 - 0,90) Chi-cuadrado = 4,10; gl = 5 (ρ = 0,5356) Inconsistencia (l^2) = 0,0 %



CP positivo (95% CI)

Mosimann 2017	7,87	(4,61 - 13,41)
O'Gorman 2017	9,72	(8,79 - 10,74)
Guizani 2018	8,34	(6,62 - 10,50)
Tan 2018		(8,42 - 10,20)
Lobo 2019		(5,83 - 12,60)
Boutin 2021	6,56	(4,34 - 9,93)

Modelo de efectos fijos CP positivo ponderado = 8,86 (8,21 - 9,56) Cochran-Q = 6,59; gl = 5 (ρ = 0,2528) Inconsistencia (l^2) = 24,2 %



CP negativo (95% CI)

Mosimann 2017	0,19	(0,01 - 2,34)
O'Gorman 2017	0,03	(0,00 - 0,47)
Guizani 2018	0,19	(0,07 - 0,52)
Tan 2018	0,11	(0,05 - 0,24)
Lobo 2019	0,16	(0,03 - 0,97)
Boutin 2021	0,34	(0,13 - 0,87)

Modelo de efectos fijos CP negativo ponderado = 0,13 (0,08 - 0,22) Cochran-Q = 5,45; gl = 5 (p= 0,3640) Inconsistencia (l^2) = 8,2 %

	AETSA — EVALUACIÓN DE TECNOLOGÍAS SANITARIAS DE ANDALUCÍA	
Junt	Ta Consejería de Salud	https://www.aetsa.orσ