

AGENCIA DE EVALUACIÓN
DE TECNOLOGÍAS SANITARIAS
DE ANDALUCÍA (AETSA)

RENDIMIENTO DIAGNÓSTICO DE LOS MODELOS DE PREDICCIÓN CLÍNICA DE EMBOLISMO PULMONAR

DIAGNOSTIC PERFORMANCE OF CLINICAL
PREDICTION MODELS FOR PULMONARY EMBOLISM
EXECUTIVE SUMMARY

CONSEJERÍA DE SALUD

Agencia de Evaluación de Tecnologías
Sanitarias de Andalucía (**AETSA**)

Informe de Evaluación

Rendimiento diagnóstico de los modelos de predicción clínica de embolismo pulmonar



Avda. de la Innovación s/n. Edificio Renta Sevilla. 2ª planta

41020 Sevilla-España (Spain)

Tlf.: 955 00 68 04 - Fax: 955 00 68 45

e-mail: aetsa.csalud@juntadeandalucia.es

Sevilla, marzo 2008

2008

INFORME_3

López Márquez R

Rendimiento diagnóstico de los modelos de predicción clínica de embolismo pulmonar . Reyes López Márquez, Soledad Márquez Calderón. Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, 2008.

72 p.; 127 cm

1. Evaluación de Tecnología Biomédica 2. Embolia Pulmonar 3. Predicción 4. Diagnóstico Clínico I. Márquez Calderón S. II. Andalucía. Agencia de Evaluación de Tecnologías Sanitarias

Agencia de Evaluación de Tecnologías Sanitarias de Andalucía

www.juntadeandalucia.es/salud/orgdep/AETSA

Autoras: Reyes López Márquez y Soledad Márquez Calderón

Conflicto de interés: Las autoras declaran no tener ningún conflicto de interés secundario (económico o personal) que pueda influir en su juicio profesional concerniente al interés primario relacionado con los objetivos de este informe.

Este documento puede citarse como:

López Márquez R, Márquez Calderón S. Rendimiento diagnóstico de los modelos de predicción clínica de embolismo pulmonar [Internet]. Sevilla: Agencia de Evaluación de Tecnologías Sanitarias, 2008. Informe 3/2008. URL: <http://www.juntadeandalucia.es/salud/AETSA>.

Este documento puede ser reproducido total o parcialmente, por cualquier medio, siempre que se mencione explícitamente la fuente.

ISBN: 978-84-691-4638-3

Depósito Legal:

Agradecimientos

Este documento se ha beneficiado de forma importante de las aportaciones de los siguientes revisores externos:

Dra. Remedios Otero Candelera. Médica Especialista en Neumología. Unidad de Gestión Clínica de Enfermedades Respiratorias. Hospitales Universitarios Virgen del Rocío, Sevilla.

Dr. Carlos Palacios Gómez. Médico Especialista en Medicina Familiar y Comunitaria. Servicio de Cuidados Críticos y Urgencias. Hospitales Universitarios Virgen del Rocío, Sevilla.

Dr. Javier Trujillo Santos. Médico Especialista en Medicina Interna. Servicio de Medicina Interna. Hospital Universitario Santa María del Rosell, Cartagena.

La Agencia de Evaluación de Tecnologías Sanitarias de Andalucía y las autoras reconocen y agradecen a los revisores su dedicación y aportaciones.

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

Índice

Resumen ejecutivo.....	8
Executive summary.....	11
Introducción.....	13
1. Epidemiología y principales características del embolismo pulmonar	13
2. El papel de la valoración clínica en el diagnóstico del embolismo pulmonar.....	14
3. Justificación de la revisión sistemática.....	17
Objetivos.....	20
Material y métodos.....	21
Tipo de estudio.....	21
Estrategia de búsqueda	21
Criterios de inclusión y exclusión.....	22
Evaluación crítica de los estudios seleccionados.....	24
Extracción de datos y resultados.....	24
Síntesis de los resultados.....	25
Resultados.....	27
Selección de artículos según criterios de inclusión	27
Descripción de los estudios seleccionados.....	29
Valoración crítica y resultados de los estudios seleccionados.....	39
Estudios que analizan una escala ó modelo predictivo sin otra prueba diagnóstica de comparación.....	39
Estudios que analizan una escala ó modelo predictivo comparándolas con el empirismo.....	44
Estudios que analizan varias escalas comparándolas entre sí	51
Síntesis de resultados sobre rendimiento diagnóstico.....	59
1. Escalas frente a empirismo.....	59
2. Rendimiento diagnóstico de las diversas escalas.....	59
2.1. Escala de Wells.....	59
2.2. Escala de Ginebra.....	60
2.3. Otras escalas.....	61
Discusión.....	63
Limitaciones y aspectos metodológicos.....	63
Discusión de los resultados principales.....	64
Implicaciones para la práctica clínica y la investigación.....	66
Conclusiones.....	69
Referencias.....	71
ANEXO I. Cuestionario QUADAS	75
ANEXO II. Exclusiones.....	76

Resumen ejecutivo

Título: Rendimiento diagnóstico de los modelos de predicción clínica de embolismo pulmonar.

Autoras: Reyes López Márquez, Soledad Márquez Calderón.

Solicitado por: Comisión de Asistencia Sanitaria basada en la Evidencia del Hospital Virgen del Rocío (Sevilla) a través del Grupo de Trabajo sobre Embolismo Pulmonar de dicho centro.

Fecha de inicio: 20 septiembre 2007.

Fecha de finalización: 29 febrero 2008.

Antecedentes y justificación: Las manifestaciones clínicas del embolismo pulmonar son muy poco sensibles y específicas, por lo que, ante la sospecha de esta patología, el clínico se maneja con un enorme grado de incertidumbre. Para facilitar el enfoque de una estrategia diagnóstica adecuada, se han desarrollado numerosas herramientas (escalas, modelos, reglas), que ayudan a determinar la probabilidad de que el paciente presente esta patología, de forma previa a la realización de pruebas complementarias invasivas que confirmen o descarten el diagnóstico de forma definitiva. Hasta el momento, no se tiene clara evidencia de que el rendimiento diagnóstico de estas escalas o modelos sea mejor que el proporcionado por la experiencia del clínico, o sobre cuál de las escalas se comporta con mayor exactitud. El rendimiento diagnóstico de estas herramientas ha sido estudiado en diversos trabajos, que son el objeto de la actual revisión.

Objetivos: El objetivo principal ha sido evaluar el rendimiento diagnóstico de las escalas ó modelos clínicos utilizados para la determinación de la probabilidad pretest en el diagnóstico del embolismo pulmonar. Partiendo de la bibliografía encontrada, se planteó como objetivo secundario describir otros resultados de las escalas ó modelos diferentes al rendimiento diagnóstico.

Metodología: Se ha realizado una revisión sistemática de la literatura. Inicialmente, se realizó una búsqueda bibliográfica exploratoria, enfocada a identificar revisiones sistemáticas, en MEDLINE, Emergency Care Research Institute, Cochrane Library, Centre for Reviews and Dissemination, International Network of Agencies for Health Technology Assessment (INAHTA) y National Guideline Clearinghouse. Posteriormente se realizó la búsqueda definitiva, dirigida a localizar estudios originales, en MEDLINE y EMBASE. Los principales criterios utilizados para la selección de los artículos se han referido a población (pacientes adultos con sospecha de embolismo pulmonar), intervención (alguna herramienta estructurada para estimar la probabilidad pretest de embolismo pulmonar), patrón oro (pruebas complementarias habituales para el diagnóstico del embolismo y/o seguimiento clínico de 3 meses), resultados (rendimiento diagnóstico y/o resultados clínicos de la aplicación de las escalas) y diseño (ensayos clínicos y estudios observacionales con ó sin grupo control). La evaluación crítica de los trabajos originales se realizó con la herramienta QUADAS. Se ha realizado una síntesis cualitativa

de los datos.

Resultados: En la búsqueda definitiva resultaron 428 referencias. Tras eliminar 38 duplicados, quedaron 390 documentos. Tras la primera selección, basada en la lectura de título y resumen, se excluyeron 315 artículos, quedando 75 para lectura a texto completo. De ellos, siguiendo los criterios de inclusión y exclusión, se excluyeron 68, quedando finalmente incluidos 11 trabajos originales en la revisión; todos ellos estudios observacionales sin grupo control. En la mayoría de los estudios se valoraba más de una escala.

La rentabilidad del uso de escalas frente al empirismo se ha estudiado en tres trabajos de buena calidad que no encontraron diferencias relevantes entre el rendimiento diagnóstico de estas dos opciones de cálculo de la probabilidad pre-test de embolismo pulmonar.

Cuando se comparó la rentabilidad de las diferentes escalas entre sí, los resultados presentaron bastante variabilidad. La escala de Wells de 2 categorías se ha mostrado más específica que sensible en dos estudios de buena calidad; en ambos la sensibilidad osciló entre el 60 y 62%. La escala de Wells de 3 categorías ha sido evaluada en 9 estudios. Dos estudios la valoraron en poblaciones con baja frecuencia de embolismo pulmonar, aportando resultados dispares en sensibilidad y especificidad (en un trabajo se estimó una sensibilidad del 92% y una especificidad moderada, y en el otro la sensibilidad fue del 54%), aunque ambos coincidieron en el alto valor predictivo negativo de la escala (superior al 96%). Con frecuencias moderadas y altas de embolismo pulmonar los estudios coincidieron en presentar la escala de Wells de tres categorías como más sensible que específica: con frecuencias moderadas la sensibilidad fue superior al 90% (dos estudios), y con frecuencias altas (cinco estudios), ésta osciló entre el 66 y el 91%. La escala de Ginebra, valorada en cinco estudios, se caracterizó por ser más sensible que específica. En el único estudio en el que esta escala fue aplicada de forma prospectiva, sus resultados compararon favorablemente con los de la escala de Wells, ambos con sensibilidad superior al 70%. El rendimiento diagnóstico del algoritmo de Wells, la regla de Charlotte y los modelos de Rodger y Pisa ha sido evaluado en un solo estudio para cada uno de ellos. De todos ellos, sólo el modelo de Pisa ha demostrado una rentabilidad diagnóstica superior, de forma estadísticamente significativa, a la de las escalas de Wells y Ginebra (área bajo la curva ROC del 94% para este modelo, frente al 54% de la escala de Ginebra y 75% de la escala de Wells). Aunque se trata de un trabajo de buena calidad, los problemas de validez externa limitan su aplicabilidad. En estos estudios no se han registrado resultados de efectividad clínica de la aplicación de las escalas.

Conclusiones: La rentabilidad diagnóstica de las escalas usadas para la determinación de la probabilidad pretest en el diagnóstico del embolismo pulmonar no presenta diferencias relevantes con la rentabilidad proporcionada por la valoración empírica. Con la evidencia disponible no se ha podido poner de manifiesto la superioridad de la rentabilidad diagnóstica de una escala concreta sobre las demás. Las escalas de Wells de tres categorías y Ginebra son las más estudiadas. La escala de Wells de tres categorías ha mostrado una alta sensibilidad (mayor al 90%) en poblaciones con moderada frecuencia de embolismo pulmonar; pero en poblaciones con alta frecuencia los

datos de sensibilidad fueron discordantes (entre 66 y 91%). En estudios de buena calidad en los que se ha evaluado la rentabilidad de la escala de Ginebra en poblaciones de alta frecuencia de embolismo pulmonar la sensibilidad ha resultado ser alta (mayor de 86%). Los estudios sobre rentabilidad diagnóstica hallados en esta revisión no han valorado la efectividad clínica de las escalas en términos de resultados en salud.

Executive summary

Title: Diagnostic performance of clinical prediction models for pulmonary embolism.

Authors: Reyes López Márquez, Soledad Márquez Calderón.

Requested by: Evidence-based Health Care Commission of Virgen del Rocío Hospital (Seville) through the Pulmonary Embolism Work Team of the said centre.

Start date: September 20th 2007.

Completion date: February 29th 2008.

Background And Rationale: The clinical manifestations of pulmonary embolism are just a bit sensitive and specific, so that, the clinician has to face a considerable amount of uncertainty when suspecting this pathology. To facilitate the approach to an appropriate diagnostic strategy, there have been developed several tools (scores, models, rules) which help to determine the patient's probability to have the pathology prior to performing invasive supplementary tests that definitely confirm or discard the diagnostic of embolism. Up to the moment, there is no clear evidence proving that the diagnostic performance of these scores or models is better than that provided by the clinician's experience or reporting which of the scores is more accurate. The diagnostic performance of these tools has been studied in several papers, which are the subject of the current review.

Objectives: The main objective has been to assess the diagnostic performance of the scores or clinical models used to determine the pre-test probability in diagnosing pulmonary embolism. A secondary objective was to describe other outcomes from the scores or models, different to the diagnostic performance.

Methodology: A literary systematic review was conducted. Initially, an exploratory bibliographic search was run which was addressed to identify systematic reviews on MEDLINE, Emergency Care Research Institute, Cochrane Library, Centre for Reviews and Dissemination, International Network of Agencies for Health Technology Assessment (INAHTA) and National Guideline Clearinghouse. Afterwards, the definite search was conducted which was focused on finding original studies on MEDLINE and EMBASE. The main criteria used to select the papers have referred to patient population (adults with suspected pulmonary embolism), intervention (any tool which has been structured to estimate pre-test probability of pulmonary embolism), gold standard (usual and supplementary tests to diagnose embolism and/or 3 months clinical follow-up), results (diagnostic performance and/or clinical outcomes from applying the scores) and design (clinical trials and observational studies with or without control group). The critical assessment of the original studies was conducted with QUADAS tool. The data were summarised qualitatively.

Results: The definite search resulted in 428 references. 390 documents remained after discarding 38 duplicates. After the first selection was completed, which had been based on reading title and summary, 315 papers were excluded, so 75 remained to be read on its whole. Of the latter, 68 papers were excluded following the inclusion and exclusion

criteria so that 11 original studies were finally included in the review. All of them were observational studies without control group. Most of them assessed more than one score.

It has been studied the accuracy of the scores as opposed to empiricism in three, good quality works. The studies did not find any relevant difference in the diagnostic performances of these two options when estimating the pre-test probability of pulmonary embolism.

When accuracy of the scores was compared with each other, the results were much variable. 2 categories-Wells score has proved to be more specific than sensitive in two good quality studies; sensitivity varied in both of them from 60% to 62%. 3 categories-Wells score has been assessed in 9 studies. Two studies assessed it in groups of patients with low frequency of pulmonary embolism, providing different results on sensitivity and specificity (it was estimated 92% of sensitivity and moderate specificity in a study while another presented 54% of sensitivity), although both studies coincided in high negative predictive value of the score (over 96%). In patients with moderate and high frequency of pulmonary embolism, the studies matched in presenting the 3 categories-Wells score as more sensitive than specific: sensitivity was over 90% in patients with moderate frequency of embolism (two studies), and it varied between 66 and 91% (five studies) in patients with high frequency of embolism. The Geneva score, which was assessed in five studies, was characterised by being more sensitive than specific. Its results matched those of Wells scores favourably in the only one study that had applied the score prospectively. Both results achieved over 70% of sensitivity. The diagnostic performance of Wells algorithm, Charlotte rules and Rodger and Pisa models have been assessed each of them in one study, only. Of all them, just the Pisa model has proved, in a statistically significant way, a superior diagnostic performance than Wells and Geneva scores (94% of the area under the ROC curve for this model as opposed to 54% for the Geneva score and 75% for Wells score). Although it is a good quality study, problems related to external validity jeopardise its applicability. It has been reported no result on clinical effectiveness of the scores application in these studies.

Conclusions: The diagnostic performance of the scores used to determine the pre-test probability in diagnosing pulmonary embolism does not present relevant differences with the diagnostic performance provided by the empirical assessment. On the basis of the evidence available, there cannot be established the superiority of the diagnostic performance of a specific score over the rest of them. 3 categories-Wells score and Geneva score have been studied the most. 3 categories-Wells score has proved high sensitivity (higher than 90%) in patients with moderate frequency of pulmonary embolism, but in patients with high frequency, sensitivity data did not match (between 66 and 91%). In good quality papers in which it has been assessed the accuracy of Geneva score in patients with high frequency of pulmonary embolism, sensitivity has proved to be high (over 86%). The studies on diagnostic performance found in this review have not assessed clinical effectiveness of the said scores in terms of health outcomes

Introducción

> 1. Epidemiología y principales características del embolismo pulmonar

El embolismo pulmonar (EP) se origina generalmente en el contexto de una trombosis venosa profunda (TVP); ocurre cuando un trombo originado en una vena, normalmente en el sistema venoso profundo de las extremidades inferiores, se desplaza al árbol arterial pulmonar (1). El EP es una entidad potencialmente mortal con un rango de presentaciones clínicas que oscila entre la inestabilidad hemodinámica y la ausencia de síntomas (2). Constituye la manifestación más seria de la enfermedad tromboembólica venosa (1), y una complicación de numerosas y diferentes entidades (2).

Más que de una enfermedad bien definida, se trata de un síndrome cuyo diagnóstico es a menudo difícil de obtener y es frecuentemente obviado (2), por lo que resulta muy difícil conocer su verdadera incidencia. Se ha estimado que en un hospital general con una población de referencia de 200.000 habitantes se podrían diagnosticar unos 50 casos de EP al año; pero se estima que la verdadera incidencia del EP es probablemente superior al 1% de todas las admisiones al hospital, ya que muchos de estos casos sólo se harán evidentes en los estudios necrópsicos (3). A nivel de la comunidad, se han realizado grandes estudios en los que se estima que la incidencia global anual es de 60-70 casos de EP por 100.000 habitantes (4). Aún así, no puede descartarse que la incidencia sea más alta, considerando que el número de casos de EP asintomáticos no fatales no puede ser determinado (2).

1.1. Fisiopatología del embolismo pulmonar. Factores de riesgo

El EP y la TVP deben considerarse parte de un mismo proceso fisiopatológico (5). La patogénesis de la enfermedad tromboembólica venosa, la condición que causa el EP, se basa en la tríada de Virchow: éstasis circulatorio, daño en la pared vascular e hipercoagulabilidad. Estas tres condiciones se presentan conjuntamente en determinadas situaciones adquiridas ó heredadas que denominamos factores de riesgo (2).

Aunque el EP puede ocurrir en sujetos que no presentan ningún factor de riesgo (2), en todos los estudios de EP se encuentran uno ó más factores predisponentes en el 80-90% de los pacientes (3). La predisposición genética (deficiencia hereditaria de factores que inhiben o regulan los procesos de coagulación ó fibrinólisis) sólo explica una quinta parte de los casos de EP, correspondiendo el grupo más amplio a los pacientes con condiciones adquiridas favorecedoras de EP (5).

Entre los factores de riesgo más frecuentemente relacionados con el desarrollo del EP se encuentran la inmovilización prolongada, el antecedente de enfermedad tromboembólica, cirugía o fracturas recientes (sobre todo en miembros inferiores) y el cáncer (3,5). El papel de la edad como factor de riesgo independiente no está tan claro pues, aunque la incidencia de EP aumenta exponencialmente con la edad, ello puede ser reflejo de la mayor prevalencia de enfermedades e intervenciones quirúrgicas en edades

mayores de la vida (3).

Debe reconocerse que el riesgo de EP aumenta con el número de factores de riesgo presentes. De hecho, la presencia ó ausencia de factores de riesgo para la enfermedad tromboembólica venosa es esencial en la evaluación de la probabilidad de EP (2).

1.2. Características clínicas

El espectro del EP varía desde un embolismo clínicamente insignificante hasta un embolismo masivo con muerte súbita, dependiendo del tamaño del émbolo y de la reserva cardiorrespiratoria del paciente (5). Dentro de este rango, las posibilidades de presentación de manifestaciones clínicas, solas o en combinación, son múltiples y muy variadas. Entre los síntomas más frecuentemente referidos se encuentran la disnea, dolor pleurítico, tos, aumento del perímetro del miembro inferior, hemoptisis y palpitaciones. En la exploración física los signos más frecuentes son taquipnea, estertores, taquicardia, cuarto tono, refuerzo del componente pulmonar del segundo tono (5), e incluso signos más inespecíficos como la fiebre.

Se trata, en definitiva, de manifestaciones clínicas comunes a otros procesos o enfermedades respiratorias o cardiológicas, por lo que se plantea un amplio espectro de diagnósticos diferenciales.

1.3. Pronóstico

El EP puede ocurrir de forma aislada o en forma de eventos sucesivos. Tras un episodio de EP hay un riesgo importante de recurrencia, especialmente durante las primeras 4-6 semanas. Este riesgo aumenta considerablemente en ausencia de terapia anticoagulante (2). En el estudio multicéntrico Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) (6,7), la mortalidad global a los 3 meses fue del 15%. El 75-90% de los fallecimientos tendría lugar en las primeras horas de producirse el EP. En el 10-25% restante, la causa de muerte probablemente sean los embolismos recurrentes que tienen lugar en las dos semanas siguientes (5). La mortalidad del EP no tratado se estima en torno al 20-30% (2).

> 2. El papel de la valoración clínica en el diagnóstico del embolismo pulmonar

En la práctica clínica, pueden ocurrir al mismo tiempo dos fenómenos relacionados con los errores diagnósticos del EP: éste puede no diagnosticarse en algunos casos (infradiagnóstico) o ser diagnosticado en personas que no lo presentan (sobrediagnóstico). El infradiagnóstico lleva a un grupo de pacientes a no recibir el tratamiento adecuado para un problema potencialmente mortal. Según la Task Force on Pulmonary Embolism (European Society of Cardiology), un meta-análisis de 12 estudios post-mortem llevados a cabo entre 1971 y 1995 mostró que más del 70% de los EP mayores no habían sido sospechados por el clínico (2). En el otro extremo, con el sobrediagnóstico, se expone de forma innecesaria a un grupo de pacientes a un tratamiento, la terapia anticoagulante, que

puede comprometer la vida (3).

La sospecha clínica de EP se establece en base a los signos y síntomas iniciales, que son considerados a la luz de la presencia o ausencia de factores de riesgo (1). Sin embargo, la valoración clínica en los pacientes con sospecha de EP es notoria por su inexactitud, debida a la falta de sensibilidad y especificidad de los síntomas, signos y hallazgos de las pruebas rutinarias, como el electrocardiograma, la gasometría arterial, o la radiografía de tórax (8). Aunque la presencia de más de un signo clínico ó síntoma aumenta la sensibilidad (1), el valor predictivo global de cualquiera de estos hallazgos por separado para el diagnóstico o la exclusión del EP es inferior al 80% (3).

Pese a estas premisas, aparentemente decepcionantes, sobre la utilidad de la valoración clínica, los resultados de estudios prospectivos a gran escala sobre el diagnóstico del EP prestan soporte a la idea de que la valoración clínica es un paso fundamental en el desarrollo del diagnóstico (9). Así, el estudio PIOPED documentó que, de los pacientes clasificados por la gammagrafía de ventilación-perfusión como de baja probabilidad de EP, un 14% presentaba EP en la angiografía; sin embargo, al subdividir a estos pacientes en dos grupos según la valoración clínica del riesgo, se observó que entre los clasificados como de bajo riesgo clínico, la frecuencia de EP fue del 4%, mientras que entre los clasificados como de alto riesgo, un 40% tuvo realmente un EP (6, 7, 10).

Una vez que los clínicos contemplan la posibilidad del EP, pueden definir la probabilidad clínica de este evento como una probabilidad pretest (PPT), que más que diagnosticar ó excluir definitivamente el EP, permite categorizar a los pacientes en subgrupos, tales como los de probabilidad baja, media o alta, con orden creciente de frecuencia de diagnóstico de EP (11). Esta probabilidad puede ser utilizada en la selección de pacientes para la realización de posteriores pruebas diagnósticas (9). Así, cuando la probabilidad clínica (PPT) se combina adecuadamente con datos suplementarios, permite obviar las pruebas de imagen en una significativa proporción de pacientes, minimiza la necesidad de pruebas invasivas y optimiza el coste-efectividad (12). De esta forma, evaluar la probabilidad de EP en el paciente individual de acuerdo a la presentación clínica es de la máxima importancia en la selección de una estrategia de diagnóstico adecuada y en la interpretación de los resultados de las pruebas diagnósticas (2).

Hasta hace poco, la agrupación de pacientes en categorías de baja, media o alta probabilidad de EP ó valoración de la PPT era enteramente implícita, y los médicos asumían que sus propias estimaciones de la probabilidad clínica eran lo suficientemente válidas. Actualmente existen nuevas herramientas que proporcionan al clínico criterios más explícitos para estimar la probabilidad clínica de EP (13).

2.1. Valoración mediante el empirismo

El intrincado proceso mediante el que el clínico formula una hipótesis sobre el origen de la enfermedad es difícil de explicar y de enseñar, porque ocurre en el pensamiento humano y es dependiente de los conocimientos y experiencias previos (14). En la literatura anglosajona se emplea el término “gestalt” (derivado de una palabra alemana), que se podría definir como una pauta, modelo ó estructura completa que es percibida como algo mayor que la simple suma de sus partes separadas. En Medicina, estas partes separadas son fragmentos relevantes de información obtenida de la historia médica y del examen

físico que son asimiladas por el clínico (14) y analizadas a la luz de su conocimiento y experiencia. En castellano, aunque no existe traducción para este término, utilizamos, en este mismo sentido, los términos juicio clínico o empirismo (ó más coloquialmente “ojo clínico”).

En la práctica diaria muchos clínicos eligen no utilizar algoritmos estructurados para valorar la PPT y en vez de esto confiar en su juicio clínico (14). Pese a que anteriores estudios han puesto de manifiesto la escasa habilidad de los clínicos, fundamentalmente de aquellos que cuentan con menor grado de experiencia, para asignar una probabilidad clínica (3), el efecto del nivel de entrenamiento, y la exactitud y reproducibilidad de esta estimación no estructurada continua siendo un tema controvertido (14).

2.2. Valoración mediante modelos de predicción clínica

En los últimos años, muchos investigadores han desarrollado y puesto a prueba varios modelos de decisión estructurados para la estimación de la PPT, entre los que se incluyen sistemas de puntuación y algoritmos (14).

En su primer trabajo, publicado en 1998, Wells y colaboradores propusieron un algoritmo que resultó de un proceso de consenso (15). Pese a sus limitaciones a la hora de utilizarlo en la práctica clínica, dada su complejidad, este trabajo inicial fue importante porque representó la formalización del proceso de predicción clínica a través de unos criterios explícitos (13).

Los propios autores desarrollaron y publicaron, en el año 2000, un modelo revisado (16), más fácil de recordar y aplicar. Mediante un procedimiento de regresión logística se determinaron los predictores clínicos que se asociaban de forma independiente con la presencia de EP. Resultaron asociarse 7 variables, que constituyen los elementos de esta escala: signos y síntomas de TVP, frecuencia cardíaca superior a 100 latidos por minuto, inmovilización o cirugía en las 4 semanas previas, antecedentes personales de TVP ó EP, hemoptisis, diagnóstico de un proceso maligno y la consideración de que un diagnóstico alternativo sea menos probable que el EP. El resultado es una puntuación global, que se obtiene tras sumar las puntuaciones asignadas a cada variable según el coeficiente de regresión estimado. Esta puntuación global puede ser resumida de dos formas, según si la PPT se divide en dos o tres categorías. Así, la escala de Wells con 3 categorías separa a los pacientes según si la PPT es baja (menos de 2 puntos), intermedia (2-6 puntos) y alta (más de 6 puntos). La escala de Wells con dos categorías ó dicotomizada los clasifica en pacientes con EP poco probable (4 puntos o menos) y con EP probable (más de 4 puntos) (13).

En 2001, Wicki y colaboradores publicaron un nuevo modelo, conocido como escala de Ginebra (17). Siguiendo un procedimiento de regresión logística, se identificaron 7 factores asociados independientemente con la presencia de EP: edad, antecedentes personales de EP ó TVP, cirugía reciente, frecuencia cardíaca superior a 100 latidos por minuto, descenso en la presión arterial de dióxido de carbono (por debajo de 39 mmHg), descenso en la presión arterial de oxígeno (por debajo de 82,4 mmHg), y anomalías específicas en la radiografía de tórax. Como ocurría con la escala de Wells, la puntuación final alcanzada es categorizada en tres posibles valores, que corresponden a las categorías de PPT baja, intermedia y alta (puntuación menor a 5, de 5 a 8, igual ó

superior a 9, respectivamente) (13).

Otras escalas ó modelos han sido menos estudiadas, siendo por tanto menos conocidas; entre otros destacan la regla de Charlotte (18), el modelo de Pisa (9) y el modelo de Rodger (19).

La regla de Charlotte (18) es un modelo de 6 variables con forma de algoritmo, desarrollado a partir de un análisis estadístico con regresión logística. El modelo utiliza inicialmente dos variables (la edad y el índice de shock) como cribado; y si éstas presentan un determinado valor, se continúa aplicando el modelo. Las otras 4 variables son: descenso en la presión arterial de oxígeno (no explicada por otro proceso), tumefacción unilateral en miembro inferior, cirugía reciente y hemoptisis. El modelo clasifica a los pacientes en dos categorías, con respecto a su PPT de EP: no-alto riesgo y alto riesgo.

El grupo de Miniati y colaboradores desarrolló y validó un modelo de 15 variables (modelo de Pisa) (9), entre las que se incluían: sexo, edad, enfermedad preexistente (2 variables), presencia de tromboflebitis, presencia de síntomas de EP (3 variables), fiebre, signos electrocardiográficos, y alteraciones en la radiografía de tórax (5 variables). A diferencia de otros modelos estructurados, en éste los autores utilizan directamente los valores de los coeficientes de correlación para calcular la PPT, mostrando así una estimación real para cada paciente individual. Tras convertir esta estimación en un porcentaje, la PPT es categorizada en uno de cuatro posibles valores: baja, moderada, moderadamente alta y muy alta.

El modelo de Rodger (19), publicado en 1999, excluye el EP si todas las variables siguientes están ausentes: dímero D semicuantitativo positivo, frecuencia cardíaca superior a 100 latidos por minuto, dolor o tumefacción en miembro inferior, eventos tromboembólicos previos y cirugía reciente. Considera, entonces, dos categorías de PPT: EP probable y EP no probable.

En definitiva, todos estos modelos y escalas son herramientas, casi siempre derivadas de análisis estadísticos, que ofrecen estimaciones de la probabilidad de enfermedad sobre la base de un conjunto de datos clínicos. Con ellas se pretende estandarizar y representar de forma explícita la valoración de la probabilidad pretest de EP (12). Las características, exactitud, validación y reproducibilidad de estos modelos han sido estudiadas y revisadas en diversos trabajos, en los que se comparan con el empirismo o diversas escalas entre sí (1).

> 3. Justificación de la revisión sistemática

Esta revisión surge a propuesta de un grupo de trabajo multidisciplinar sobre tromboembolismo pulmonar constituido en los Hospitales Universitarios Virgen del Rocío. Se trata de la primera pregunta de investigación de una serie de preguntas sucesivas con las que se pretende abarcar todo el espectro diagnóstico del EP: desde el análisis del mejor método para estimar la probabilidad pretest hasta la determinación del patrón oro

para la confirmación diagnóstica del EP, pasando por el estudio del rendimiento de otras pruebas complementarias utilizadas en el diagnóstico del EP como el dímero D, la ultrasonografía de miembros inferiores, ó la gammagrafía de ventilación-perfusión. El objetivo final de esta línea de trabajo es la creación de un algoritmo diagnóstico, basado en la mejor evidencia disponible, con el que los clínicos puedan guiarse en la toma de decisiones ante la sospecha de EP en un paciente concreto.

Para comprobar el estado actual del conocimiento sobre la validez y utilidad de las escalas para valoración de la probabilidad pretest en el EP, se realizó una búsqueda bibliográfica exploratoria de forma previa a la realización de la revisión sistemática propiamente dicha. En esta búsqueda se encontraron 3 revisiones sistemáticas en las que se comparaban escalas de predicción clínica con el empirismo o escalas entre sí (11, 20, 21).

La revisión de Chunilal et al (11) se publicó en 2003 y su búsqueda bibliográfica, realizada en MEDLINE, abarcó desde 1966 a marzo de 2003. Su objetivo era evaluar el rendimiento de la valoración de la probabilidad pretest en el EP comparando el uso de escalas con el empirismo. Concluyeron que las escalas de predicción clínica obtenían un rendimiento diagnóstico similar al alcanzado utilizando el juicio clínico para los pacientes asignados a las categorías de baja y alta probabilidad. A pesar de no hallar diferencias entre empirismo y escalas, acaban recomendando el uso de éstas, con el argumento de que son simples y mantienen su rendimiento diagnóstico cuando son utilizadas por médicos menos experimentados. Sin embargo, estos autores no se decantan por el uso de una determinada escala, al no encontrar diferencias significativas en cuanto a su rendimiento.

Las otras 2 revisiones halladas (20, 21) pertenecen al mismo grupo de investigadores: Tamariz, Segal, Eng y colaboradores. El objetivo de la primera (20), publicada en 2004, era analizar el valor predictivo de las escalas en el diagnóstico del tromboembolismo venoso. Para ello realizaron una búsqueda en MEDLINE, la Cochrane Database of Systematic Reviews y el registro de la Cochrane Controlled Trials, con límite en enero de 2003. Concluyeron que el uso de las escalas en el diagnóstico del tromboembolismo venoso tiene ventajas al proporcionar un método estandarizado de valoración del riesgo de EP que los clínicos pueden utilizar para interpretar los resultados de las subsiguientes pruebas de laboratorio y radiológicas. Al ser la escala de Wells la más estudiada, emiten conclusiones sobre su utilidad en la práctica clínica: a los pacientes en los que la PPT determinada por la escala es baja y presentan un resultado negativo en la determinación del dímero D se les puede descartar la enfermedad sin necesidad de realizar más pruebas.

La segunda revisión de este grupo de investigadores, publicada a principios de 2007 (21), responde de forma secuencial a varias preguntas de investigación sobre el diagnóstico del EP. La primera de estas cuestiones se refería al valor de las escalas de predicción clínica en el diagnóstico del TVP y el EP. Para ello utilizaron las mismas bases de datos que en la revisión anterior, además de EMBASE, y extendieron la búsqueda hasta noviembre de 2004. Concluyeron que sus resultados aportaban una fuerte evidencia para sustentar el uso de las escalas de predicción para establecer la probabilidad pretest,

sin pronunciarse sobre las potenciales ventajas de utilizar alguna de las escalas en particular. Sin embargo, esta conclusión de los autores no puede considerarse bien fundamentada en los trabajos que incluyen en la revisión.

Aunque todas estas revisiones acaban recomendando el uso de escalas para el cálculo de la PPT de embolismo pulmonar, se pueden encontrar algunas controversias entre ellas en cuanto a la evidencia existente (o la interpretación que los autores hacen de ella) en términos del rendimiento diagnóstico de las escalas comparadas con el empirismo. Así, mientras que Chunilal et al no encuentran diferencias entre ambas formas de calcular la PPT (11), Tamariz et al concluyen que hay fuerte evidencia para apoyar el cálculo de la PPT mediante escalas (21). Por otra parte, la búsqueda más reciente de estas revisiones cubre hasta el año 2004, y es posible que se hayan producido nuevas evidencias en los siguientes años, dado que es un tema de actualidad.

Todas estas razones llevan a la conclusión de que sería conveniente revisar la literatura en busca de nuevas evidencias sobre el rendimiento diagnóstico de los diferentes modelos o escalas, para determinar si su uso aporta ventajas con respecto al utilización del empirismo en la determinación de la PPT, y en caso afirmativo, determinar si alguno de estos modelos resulta más rentable. Adicionalmente, tendría gran interés matizar los resultados considerando diferentes aspectos que pueden influenciar el proceso de diagnóstico del EP: el ámbito clínico en el que aparece la sospecha de EP (departamentos de urgencias, unidades de hospitalización), la procedencia de los pacientes (si se trata de pacientes hospitalizados ó ambulatorios), o la experiencia de los médicos que atienden los casos.

Objetivos

Principal

Evaluar el rendimiento diagnóstico de las escalas ó modelos clínicos utilizados para la determinación de la probabilidad pretest en el diagnóstico del embolismo pulmonar.

Secundario

Partiendo de la bibliografía encontrada en la revisión realizada para dar respuesta al objetivo principal se planteó el siguiente objetivo secundario:

Describir otros resultados de las escalas ó modelos, diferentes al rendimiento diagnóstico.

Material y métodos

> Tipo de estudio

Revisión sistemática de la literatura médica con el objeto de identificar estudios originales y revisiones sistemáticas publicados entre 1996 y 2007.

> Estrategia de búsqueda

Bases de datos

De forma previa a la realización de la revisión sistemática propiamente dicha se realizó una búsqueda bibliográfica exploratoria para comprobar el estado actual del conocimiento sobre la validez de las escalas en la valoración de la probabilidad pretest en el embolismo pulmonar (EP), así como la pertinencia de realizar una revisión sistemática sobre el tema. Para ello se consultaron las principales fuentes de información sobre revisiones de la literatura publicadas ó en curso y sobre guías de práctica clínica: Emergency Care Research Institute (ECRI), Cochrane Library en las versiones en inglés y castellano, Centre for Reviews and Dissemination (CRD) de la Universidad de York, International Network of Agencies for Health Technology Assessment (INAHTA), la National Guideline Clearinghouse (NGC) y MEDLINE. Sólo se encontraron dos revisiones sistemáticas en MEDLINE (11, 21); si bien, a partir del listado de bibliografía de una de ellas (21) se tuvo conocimiento de una tercera revisión (20).

Posteriormente se realizó la búsqueda definitiva, dirigida a localizar estudios originales, en MEDLINE y EMBASE.

Estrategias de búsqueda

Búsqueda en **MEDLINE** (1996-2007), realizada el 16/10/2007:

- #1 (Wells OR Geneva OR Miniati OR Wicki OR Rodger).m_titl.
- #2 (scale\$ OR rule\$ OR model\$ OR scor\$ OR pretest OR "pre-test").m_titl.
- #3 ((pulmonary OR lung) adj3 emboli\$).m_titl.
- #4 (1 OR 2)AND 3.
- #5 *Pulmonary Embolism/mo, di, dt, ra, ec, ri, ep, rh, th, us [Mortality, Diagnosis, Drug Therapy, Radiography, Economics, Radionuclide Imaging, Epidemiology, Rehabilitation, Therapy, Ultrasonography]
- #6 exp Decision Support Techniques/
- #7 Triage/
- #8 exp Patient Selection/
- #9 exp "Severity of Illness Index"/
- #10 6 OR 7 OR 8 OR 9
- #11 exp "sensitivity and specificity"/
- #12 5 AND 10 AND 11
- #13 12 OR 4

Búsqueda en **EMBASE** (1996-2007), realizada el 22/10/2007:

- #1 'pulmonary *3 embolism': ab, ti AND [1996-2007]/py
- #2 'lung *3 embolism': ab, ti AND [1996-2007]/py
- #3 1 OR 2
- #4 score*: ab, ti OR scale*: ab, ti OR tod*: ab, ti OR rule*: ab, ti OR model*: ab, ti OR (pretest: ab, ti OR 'pre-test': ab, ti) AND [1996-2007]/py
- #5 3 AND 4
- #6 sensitivity: ti, ab OR specificity: ti, ab OR 'predictive value': ti, ab OR 'likelihood ratio': ti, ab AND [1996-2007]/py
- #7 5 AND 6
- #8 'lung embolism'/exp AND 'clinical assessment'/exp AND [1996-2007]/py
- #9 'lung embolism'/exp AND ('accuracy'/exp OR 'prediction'/exp OR 'probability'/exp) AND [1996-2007]/py
- #10 'lung embolism'/exp AND 'medical decision making'/exp AND [1996-2007]/py
- #11 9 AND 10
- #12 7 OR 12

Resultados de la búsqueda

Con las citadas estrategias de búsqueda se encontraron en total 428 referencias, de las que 179 procedían de MEDLINE y 249 de EMBASE. Fueron detectados 38 duplicados, quedando para la selección 390 documentos.

> Criterios de inclusión y exclusión

Para la realización de esta revisión sistemática fueron seleccionados los estudios que cumplieran los siguientes criterios de inclusión:

Población: Pacientes adultos con sospecha de EP, independientemente de que fueran hospitalizados ó ambulatorios.

Se exceptuaron los estudios que incluían solamente a aquellos pacientes que presentaban un determinado factor de riesgo para EP, como los que se referían exclusivamente a pacientes quirúrgicos, oncológicos, embarazadas, etc. No fueron excluidos, sin embargo, los estudios realizados sobre población con edad avanzada.

Intervención: Estudios en los que se utilizó alguna herramienta con la que se pueda estimar de forma explícita la valoración de la probabilidad pretest (PPT) en el diagnóstico del EP.

Estas herramientas reciben diferentes denominaciones (escala, regla, sistema de puntuación, modelo predictivo, etc) e incluyen diferentes componentes. Las más básicas están constituidas por un listado de signos y síntomas y factores de riesgo a los que se les asigna una puntuación; las más complejas incluyen, además, la valoración del

resultado de pruebas complementarias como el electrocardiograma, la radiografía de tórax o la gasometría arterial. La puntuación obtenida permite la clasificación de la PPT en alguna de las categorías preestablecidas por la herramienta, habitualmente probabilidad alta, intermedia o baja.

Se incluyeron también aquellos estudios en los que la intervención consistió en la aplicación del modelo predictivo junto a la determinación de un parámetro biológico, siempre que los resultados del rendimiento diagnóstico del modelo predictivo fueran expresados de forma independiente¹.

Patrón oro: El resultado de la intervención fue comparado con los resultados obtenidos tras la aplicación de las técnicas consideradas “patrón oro” para el diagnóstico del EP: tomografía axial computerizada (TAC) helicoidal o angiografía.

Cuando la realización de estas pruebas se estimó inoportuna ó poco ética (por ejemplo, al determinarse una baja probabilidad pre-test), se consideró como comparación el seguimiento clínico, siempre que fuera de tres meses o superior, o los resultados de otras pruebas complementarias incluidas en los algoritmos habituales para el diagnóstico del EP.

Comparación: Fueron incluidos estudios sobre el rendimiento diagnóstico de las escalas (sin comparar con otra opción para calcular la PPT), estudios que comparan las escalas frente a la experiencia del clínico (empirismo) y estudios que comparan las escalas entre sí.

Resultados: Se incluyeron estudios sobre rendimiento diagnóstico y/o resultados clínicos (duración de la estancia, aparición de complicaciones, mortalidad, etc).

La evaluación del rendimiento diagnóstico de los modelos predictivos se debía presentar en cualquiera de los siguientes términos: sensibilidad, especificidad, área bajo la curva ROC, valor predictivo positivo, valor predictivo negativo, razón de verosimilitud positiva ó razón de verosimilitud negativa, ó en cualquier otra forma que permitiera la estimación de alguno de estos parámetros.

Diseño del estudio: Ensayos clínicos y estudios observacionales con ó sin grupo control, siempre que el tamaño muestral fuese superior a los 20 sujetos.

Se excluyeron los estudios originales en los que se presenta el desarrollo y/o validación inicial de las escalas ó herramientas para la determinación de la PPT, ya que representan una fase muy preliminar en el proceso de validación clínica de una nueva prueba diagnóstica. En esta revisión se han considerado sólo aquellos trabajos en los que la prueba diagnóstica se evalúa en un contexto de práctica clínica ó similar, en una muestra consecutiva de pacientes que refleja con el mayor grado de fidelidad el ámbito en el que será aplicada.

Finalmente, se consideraron para la revisión trabajos publicados en inglés, francés o castellano.

¹ El estudio fue incluido si aportaba datos sobre la exactitud del modelo predictivo (sensibilidad, especificidad, valor predictivo, etc) como si no se hubiese utilizado el parámetro biológico.

> Evaluación crítica de los estudios seleccionados

Para evaluar la calidad de los trabajos originales seleccionados para la revisión se utilizó la herramienta QUADAS (22), cuyo cuestionario se presenta en el ANEXO I. Se eligió esta herramienta debido a que fue diseñada específicamente para valorar los estudios de pruebas diagnósticas en el contexto de las revisiones sistemáticas de la literatura. Analiza aspectos relacionados con la validez externa e interna del estudio. Con respecto a la validez externa se realizan preguntas sobre el espectro de pacientes analizado, criterios de inclusión y exclusión, o la similar disposición de datos clínicos en el estudio y en las condiciones reales de aplicación de la prueba. En cuanto a la validez interna, las preguntas van dirigidas a analizar los siguientes aspectos: la adecuación de la selección del patrón oro (prueba diagnóstica de referencia); la posible aparición de sesgos de progresión, de verificación parcial ó de incorporación en relación con el uso de este patrón; la reproducibilidad del estudio, las condiciones de cegamiento en la interpretación de los resultados; y la presentación de datos correspondientes a resultados intermedios y pérdidas.

La evaluación fue realizada por una sola persona, que de forma previa había recibido entrenamiento para la cumplimentación de la herramienta. Por otro lado, en caso de dudas en determinados ítem, éstas se resolvieron con la participación de una segunda evaluadora con experiencia en el uso del QUADAS.

Aunque la herramienta no contempla una medida final de la calidad de los estudios, a modo de resumen se presenta la suma de los ítems en los que se ha respondido "sí", a fin de facilitar una visión global.

> Extracción de datos y resultados

La extracción de la información de los trabajos originales fue realizada por una única persona.

Los resultados sobre exactitud diagnóstica se valoraron en términos de sensibilidad, especificidad, área bajo la curva ROC, valor predictivo positivo, valor predictivo negativo, razón de verosimilitud positiva ó razón de verosimilitud negativa.

Para aquellos trabajos en los que los resultados eran expresados en forma de número ó proporción de pacientes en cada categoría de la probabilidad pretest y número ó proporción de pacientes con EP diagnosticado en cada categoría, se calcularon sensibilidad y especificidad y sus respectivos intervalos de confianza al 95% de forma secundaria. Para ello se utilizaron los datos en la forma que se representa en la Figura 1, que se muestra a continuación.

Figura 1	Cálculo de sensibilidad y especificidad de las escalas o modelos, a partir de los datos presentados en los trabajos originales.		
	Enfermedad confirmada	Enfermedad descartada	
Resultado positivo de la prueba	A Número de pacientes asignados a las categorías de PPT moderada ó alta que fueron diagnosticados de EP	B Número de pacientes asignados a las categorías de PPT moderada ó alta a los que les descartó el EP	Total de pacientes asignados a las categorías de PPT moderada ó alta
Resultado negativo de la prueba	C Número de pacientes asignados a las categorías de PPT baja que fueron diagnosticados de EP	D Número de pacientes asignados a las categorías de PPT baja a los que les descartó el EP	Total de pacientes asignados a las categorías de PPT baja
	Pacientes diagnosticados de EP	Pacientes con EP descartado	Total población analizada

EP: embolismo pulmonar; PPT: probabilidad *pretest*

> Síntesis de los resultados

No se ha planteado llevar a cabo una síntesis cuantitativa dada la heterogeneidad en las poblaciones y en las intervenciones realizadas en cada uno de los estudios. La variabilidad entre estudios en aspectos como los criterios de exclusión, la ubicación del paciente (hospitalizado o ambulatorio), la frecuencia de embolismo pulmonar en cada muestra, la forma de categorizar la probabilidad pretest ó las excepciones a la hora de aplicar las escalas, hace que una síntesis cuantitativa de los resultados no sea viable.

Por ello, se ha realizado una síntesis cualitativa de los resultados matizando las conclusiones por la calidad de los datos y por otros aspectos, como los relacionados con la población de estudio (pacientes hospitalizados ó ambulatorios) ó con la experiencia del clínico que realizó la intervención.

Resultados

> Selección de artículos según criterios de inclusión

Como ya se ha comentado en el apartado de Material y Método, tras descontar artículos duplicados, quedaron 390 artículos para realizar la selección.

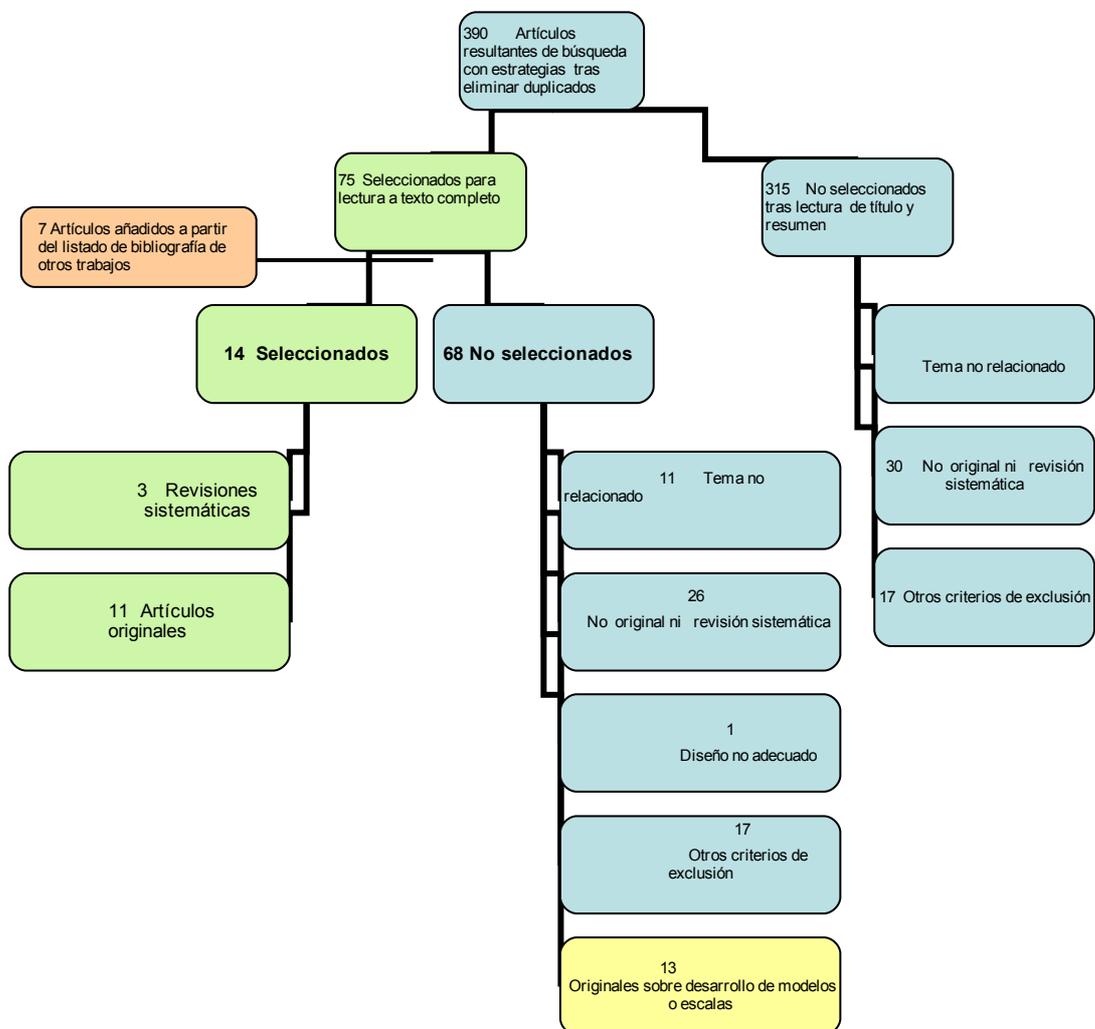
Se procedió a una primera selección de trabajos, basada en la lectura de título y resumen, tras la que fueron excluidos 315 artículos. El motivo principal de exclusión (268 de los 315 documentos) fue que el tema no estaba relacionado con la pregunta de investigación. De ellos, 113 artículos se referían fundamentalmente a modelos animales, enfermedad tromboembólica en general o trombosis venosa profunda, y otros aspectos del manejo del EP, como pronóstico, profilaxis o tratamiento. Los otros 155 artículos excluidos por el tema abordado, hacían referencia a diferentes aspectos del diagnóstico del embolismo pulmonar, entre los que no se contemplaba la valoración de la probabilidad pretest mediante el uso de escalas.

Quedaron 75 documentos para lectura a texto completo, entre los que se encontraban 2 de las 3 revisiones sistemáticas que habían sido identificadas en la búsqueda preliminar (20, 21). En los listados de la bibliografía de estas revisiones y de otros trabajos originales relevantes se localizaron 7 referencias adicionales que no habían sido halladas con las estrategias de búsqueda en MEDLINE y EMBASE, y que fueron añadidas al listado de trabajos para lectura a texto completo (entre ellas la tercera revisión sistemática (11) encontrada en la búsqueda preliminar). Así, finalmente, 82 artículos fueron seleccionados para lectura a texto completo, entre los que se incluían las tres revisiones sistemáticas.

Tras la lectura a texto completo, 68 artículos fueron excluidos de la revisión, por los motivos que se muestran detallados en el ANEXO II. Entre ellos, los principales motivos de exclusión fueron no ser estudios originales ó revisiones sistemáticas (26 estudios), y tratarse de estudios originales sobre el desarrollo y/o validación de alguna escala ó modelo (13 trabajos). En la figura 2 se resume el proceso de selección de artículos para la revisión sistemática.

Tras finalizar la selección, quedaron incluidos para la revisión 14 artículos, de los que 11 eran trabajos originales y 3 las mencionadas revisiones sistemáticas. Finalmente, se comprobó que los estudios analizados en las tres revisiones que cumplían los criterios de inclusión para este trabajo (23-25), habían sido ya incluidos. Una vez comprobado esto, la revisión sistemática se centró en los 11 artículos originales.

Figura 2 Diagrama de flujo del proceso de selección de artículos para la revisión.



> Descripción de los estudios seleccionados

Entre los 11 estudios originales seleccionados se combinan diversas formas de analizar la evidencia sobre el uso de las escalas en la determinación de la probabilidad pretest (PPT). En tres trabajos (Wells (24), Chistopher Study Investigators (26) y Miniati (27)) se utiliza una sola escala en todos los pacientes incluidos, sin que los resultados sean comparados con otras formas de estimar la probabilidad pretest. En cuatro trabajos (Sanson (23), Kline (28), Carrier (29) y Penalzoza (30)) se establece comparación entre alguna escala y el empirismo. Por último, en otros cuatro estudios (Chagnon (25), Moores (31), Miniati (32) y Ollenberger (33)), se comparan distintas escalas entre sí. Ello ha dado pie a plantearse la descripción de los estudios agrupados de esta forma.

Estudios que analizan una escala ó modelo predictivo sin otra prueba diagnóstica de comparación

Las principales características de estos trabajos (24, 26, 27) se presentan en la tabla 1.

Los tres trabajos de este grupo son estudios observacionales que parten de un grupo único de pacientes con sospecha de embolismo pulmonar (EP) e incluyen entre 425 y 3503 personas. La población fue seleccionada en el servicio de urgencias en uno de los estudios (24) y es una mezcla de pacientes hospitalizados y ambulatorios en los otros dos (26, 27), aunque la proporción de pacientes ambulatorios es muy diferente entre uno y otro (22% y 82%). Todas las poblaciones tienen en torno a un 42% de varones y la edad media oscila entre los 50 y los 68 años.

Dos estudios analizan el rendimiento diagnóstico de la escala de Wells, si bien en uno se categoriza la PPT con tres posibles valores (baja, moderada y alta) (24) y en el otro con dos (EP probable y EP poco probable) (26). El tercer trabajo analiza el modelo de Pisa y no estudia el rendimiento diagnóstico de la escala individualmente (muestra los resultados globales de un algoritmo que incluye el uso de escalas). Por ello, de este trabajo se ha tomado para la revisión sistemática sólo una parte de sus resultados; concretamente la que relaciona, en el subgrupo de pacientes con diagnóstico confirmado de EP, las diferentes categorías de la PPT con el grado de severidad de la enfermedad.

El estándar de diagnóstico es diferente para cada uno de los trabajos, empleándose en uno un algoritmo diagnóstico (24), en otro la tomografía axial computerizada (TAC) (26) y en otro la angiografía (27).

Estudios que analizan una o varias escalas comparándolas con el empirismo

En la Tabla 2 se muestra un resumen de las principales características de los cuatro estudios de este grupo (23, 28-30).

Todos son estudios observacionales que parten de un grupo único de pacientes con sospecha de embolismo pulmonar (EP). Dos de los estudios fueron llevados a cabo sobre pacientes ambulatorios (28, 30); y ambos son prospectivos². Los otros dos estudios

² En esta revisión, la calificación de prospectivo o retrospectivo para un estudio hace referencia

fueron realizados sobre una población en la que se mezclan pacientes hospitalizados y ambulatorios. De ellos, uno es un estudio sin seguimiento, con un 81% de pacientes ambulatorios (23). El otro es un estudio realizado sobre una cohorte retrospectiva (29), en el que la determinación de PPT a través del empirismo y las variables necesarias para el cálculo de escalas se habían recogido de manera prospectiva, pero el cálculo de la puntuación global de las escalas se realizó de manera retrospectiva². En este estudio no se especifica la proporción de pacientes ambulatorios.

Estos cuatro estudios incluyen un número de pacientes que oscila entre 185 y 2302. El porcentaje de varones de las muestras oscila entre el 31% y el 42%, y la edad media se sitúa entre los 45 y los 56 años. Hay que destacar que existe mucha variabilidad entre ellos en cuanto a los criterios de exclusión.

Para la valoración empírica de la PPT, sólo en dos de los estudios coinciden los puntos de corte para establecer las diferentes categorías. En el estudio de Penaloza (30) no se emplea el empirismo de forma estricta. Sólo cuando los médicos en formación tienen dudas sobre la categorización de la probabilidad utilizando la escala de Wells, consultan con su supervisor que realiza una valoración empírica, siendo el criterio de éste último el que prevalece. En los 4 trabajos se ha estudiado el rendimiento diagnóstico de la escala de Wells, y en dos de ellos se aplica ésta en las dos posibles formas que ofrece esta escala para categorizar la PPT (29, 30). El resto de escalas evaluadas son: el algoritmo de Wells (23), la regla de Charlotte (28), la escala de Ginebra y el modelo de Rodger (29).

En los 4 estudios se establece el diagnóstico de EP a través de un algoritmo, siendo éste considerado el estándar de diagnóstico en todos excepto uno (28), en el que se explicita que el estándar es el seguimiento durante 90 días. Todos expresan los resultados clave en términos de número o proporción de pacientes en cada categoría de la PPT y número o proporción de pacientes diagnosticados de EP en cada categoría, si bien en dos de ellos se presentan otros parámetros de rendimiento diagnóstico.

Estudios que analizan varias escalas comparándolas entre sí

A este grupo corresponden los cuatro estudios (25, 31, 32, 33) cuya descripción se presenta en la Tabla 3.

Todos son estudios observacionales que parten de un grupo único de pacientes. Dos de estos trabajos son estudios prospectivos: uno se realizó sobre pacientes ambulatorios del departamento de urgencias (25) y el otro, sobre una muestra compuesta por pacientes hospitalizados y ambulatorios (32). Los otros dos son estudios retrospectivos (26,31), en los que se ha analizado una población que incluye pacientes hospitalizados y ambulatorios. En uno de ellos (31) se analiza un grupo de pacientes que fueron remitidos para la realización de pruebas de imagen por sospecha de EP. En el otro (26) se incluyen dos grupos de pacientes diferenciados (uno con diagnóstico seguro de EP y otro con exclusión segura de este diagnóstico). Pero en ambos trabajos, la población de origen es

al hecho de que la intervención (aplicación de las escalas) sea realizada en el mismo ó en distinto momento al de la recogida de los datos clínicos. Cuando la aplicación de las escalas se realiza a medida que se recogen los datos necesarios para su cálculo el estudio se considera prospectivo. Si la aplicación de las escalas se realiza a posteriori, sobre datos que fueron registrados anteriormente para otro estudio ó en la historia clínica, el estudio se considera retrospectivo.

una muestra consecutiva de pacientes.

Los estudios incluyen, por término medio, unos 260 pacientes, a excepción de uno de ellos que se realizó sobre una muestra de 1487 pacientes (26). La proporción de varones es muy similar en tres de los estudios, situándose en torno al 45% (25, 26, 32). La edad media/mediana se sitúa en torno a los 55-70 años. Los criterios de exclusión son muy variables entre los cuatro trabajos.

Las escalas de Wells y Ginebra han sido analizadas en los 4 estudios. En uno de ellos (32), estas dos escalas se comparan además con el modelo de Pisa. El estándar de referencia no es homogéneo para todo el grupo: dos estudios establecen como patrón oro la angiografía (32), (26), uno presenta como estándar un algoritmo de diagnóstico (25) y para el otro el estándar está constituido por cualquiera de las pruebas complementarias con las que fue realizado el diagnóstico.

Todos presentan los resultados como número o proporción de pacientes en cada categoría de la PPT y número o proporción de pacientes diagnosticados de EP en cada categoría. Además, tres de ellos (25, 26, 32), presentan el área bajo la curva ROC correspondiente a cada escala. Uno de los estudios hace un análisis sobre el rendimiento diagnóstico de las escalas considerando las diferentes ubicaciones de los pacientes (26). Adicionalmente, el estudio de Miniati (32) presenta un resultado sobre la relación entre la categoría de PPT determinada por la escala y la severidad de la enfermedad, y en el trabajo de Moores (31) se presenta la magnitud de la relación entre la PPT determinada y la incidencia de EP en cada categoría.

Tabla 1 Principales características de los trabajos que estudian una sola escala ó modelo predictivo.

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
<p>Wells et al. 2001(24) 946 pacientes incluidos 930 pacientes analizados</p>	<p>Estudio prospecti-vo sin grupo control.</p>	<p>Cuatro centros médicos. Pacientes consecutivos con sospecha de EP: departamentos de urgencias. Criterios exclusión: menores de 18 años, sospecha de TVP en MMSS, ausencia de síntomas de EP a menos de 3 días de la presentación, tratamiento anticoagulante más de 24 horas, expectativa de vida inferior a 3 meses, contraindicación de contraste, embarazo, imposibilidad continuar seguimiento. Los pacientes analizados: 42% varones, edad media 50,5 años (DE= 18,4).</p>	<p>Estimación de la PPT a través de un modelo estructurado: la escala de Wells (16). Realizada por médicos del departamento de urgencias. Se determinó el DD en todos los pacientes (SimpliRED).</p>	<p>Estándar de referencia: algoritmo diagnóstico que incluye GVP, UC y angiografía. Además incluye la PPT y el valor del DD como elementos de decisión. El EP se consideró excluido si se determinó una baja PPT y el resultado del DD fue negativo. Todos los demás pacientes fueron incluidos en el algoritmo diagnóstico. EP presente: GVP alta probabilidad, UC anormal, angiografía anormal, presentación de un ETE durante los tres meses de seguimiento.</p>	<p>Se presenta:</p> <ol style="list-style-type: none"> 1.Nº/proporción de pacientes asignados a cada categoría de la PPT. 2.Nº/proporción de pacientes con EP en cada categoría.
<p>Christopher Study Investigators 2006 (26) 3503 pacientes incluidos 3306 pacientes analizados</p>	<p>Estudio prospecti-vo sin grupo control</p>	<p>Cinco hospitales académicos y 7 hospitales generales. Pacientes consecutivos con sospecha de EP: hospitalizados y ambulatorios. Criterios exclusión: tratamiento con heparina durante más de 24 horas, expectativa de vida inferior a 3 meses, embarazo, imposibilidad continuar seguimiento, edad inferior a 18 años, alergia a contraste intravenoso, insuficiencia renal, razones logísticas, inestabilidad hemodinámica, rehusé del consentimiento. Los pacientes analizados: 42,6% varones, edad media 53,0 años (DE= 18,4), 81,7% pacientes ambulatorios.</p>	<p>Estimación de la PPT a través de un modelo estructurado: la escala de Wells (16) con PPT categorizada en 2 valores: "EP probable" y "EP improbable". Más de 300 médicos participantes.</p>	<p>Estándar de referencia: TAC. Todos los pacientes fueron seguidos durante 3 meses. EP presente: TAC anormal, presentación de ETE en seguimiento. EP descartado: DD negativo con PPT "EP improbable", TAC normal, ausencia de ETE en el seguimiento.</p>	<p>Se presenta:</p> <ol style="list-style-type: none"> 1.Nº/proporción de pacientes asignados a cada categoría de la PPT. 2.Nº/proporción de pacientes con EP en cada categoría.

Tabla 1 Principales características de los trabajos que estudian una sola escala ó modelo predictivo. (Continuación)

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
Miniati et al. 2003 (27) 425 pacientes incluidos 390 pacientes analizados	Estudio prospectivo sin grupo control	Una institución. Para el estudio: Pacientes consecutivos con sospecha de EP: hospitalizados y ambulatorios. Criterios exclusión: incapacidad para obtener el consentimiento informado, contraindicación para angiografía pulmonar, uso de terapia anticoagulante durante más de 3 días. Para el objetivo que interesa a esta revisión: la población está constituida exclusivamente por aquellos pacientes que fueron diagnosticados de EP. Los pacientes analizados: 42% varones, edad mediana 68,5 años (rango: 15-92), 22% pacientes ambulatorios.	Estimación de la PPT a través de un modelo estructurado: el modelo de Miniati ó modelo de Pisa 104. Realizada por neumólogos. Se realizó GVP a todos los pacientes.	Estándar de referencia: angiografía pulmonar. Todos los pacientes fueron seguidos durante 1 año. EP presente: GVP anormal sugestiva de EP y PPT alta o moderadamente alta, angiografía anormal, presentación de ETE en seguimiento. EP descartado: GVP normal o casi normal, GVP anormal no sugestiva de EP y baja PPT, ausencia de ETE en el seguimiento.	En pacientes con EP se ha estimado la extensión de la obstrucción vascular en la gammagrafía como índice de la severidad de la enfermedad y relacionado ésta con las 4 categorías de la PPT determinadas por la escala.

DD= dímero D; DE=desviación estándar; EP=embolismo pulmonar; ETE= evento tromboembólico; GP= gammagrafía perfusión; GV= gammagrafía ventilación; GVP= gammagrafía ventilación/perfusión; MMII= miembros inferiores; MMSS= miembros superiores; TAC= tomografía axial computerizada; TVP= trombosis venosa profunda; UC=ultrasonografía compresión.

Tabla 2 Principales características de los trabajos que estudian una o varias escalas comparándolas con el empirismo.

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
<p>Sanson et al. 2000 (23)</p> <p>1162 pacientes incluidos 517 pacientes analizados</p>	<p>Estudio observacional sin seguimiento sin grupo control.</p>	<p>Seis hospitales docentes. Pacientes consecutivos con sospecha de EP: hospitalizados y ambulatorios. Criterios exclusión: menores de 18 años, embarazadas, indicación de terapia trombolítica, estudio previo de ETEV, imposibilidad de completar el protocolo diagnóstico en las 48 horas siguientes a la presentación. Los pacientes analizados: 42% varones, edad media 51 años (DE=18), 81% pacientes ambulatorios.</p>	<p>Estimación empírica de la PPT asignando una puntuación 0-100% en escala analógica visual. La PPT estimada fue dividida en 3 categorías: menor del 20%, 20-80% y mayor del 80%. Estimación de la PPT a través de dos modelos estructurados: el algoritmo de Wells (15) y la escala de Wells (16). <i>El algoritmo de Wells incluye una pequeña modificación: no se consideran datos de saturación de oxígeno en sangre y datos electrocardiográficos, que sí son tenidos en cuenta en el modelo original.</i></p>	<p>Estándar de referencia: algoritmo diagnóstico que incluye GP y GV, angiografía y angioTAC. Todos los pacientes entran en el algoritmo diagnóstico. EP presente: GVP de alta probabilidad ó angiografía anormal. EP descartado: GVP normal ó angiografía normal.</p>	<p>Para cada una de 3 formas de estimar la PPT se ha determinado:</p> <ol style="list-style-type: none"> 1. N°/proporción de pacientes asignados a cada categoría de la PPT. 2. N°/proporción de pacientes con EP en cada categoría. 3. Razones de probabilidad (LR) para cada categoría.
<p>Kline et al. 2006 (28)</p> <p>2302 pacientes analizados</p>	<p>Estudio prospectivo sin grupo control</p>	<p>Hospital docente. Pacientes de todas las edades con sospecha de EP: departamento de urgencias. Criterios exclusión: inestabilidad hemodinámica, incapacidad para obtener una muestra sanguínea para la determinación del DD, rehusé del consentimiento. Los pacientes analizados: 31% varones, edad media 44,7 años (DE= 15,8).</p>	<p>Estimación empírica de la PPT asignando la PPT a una de las siguientes categorías: menor del 15%, 15-40% y mayor del 40%. Estimación de la PPT a través de dos modelos estructurados: la escala de Wells (16) y la regla de Charlotte (18).</p>	<p>Estándar de referencia: resultado del seguimiento durante 90 días. Se determinó el DD en todos los pacientes. EP presente: diagnóstico de EP en la evaluación inicial (algoritmo de pruebas complementarias) ó verificación de un episodio de EP ó TVP que requirió tratamiento ó muerte por EP en los 90 días posteriores a la determinación del DD. EP descartado: pacientes vivos sin diagnóstico de EP en los 90 días siguientes al inicio del seguimiento.</p>	<p>Para cada una de 3 formas de estimar la PPT se ha determinado:</p> <ol style="list-style-type: none"> 1. N°/proporción de pacientes asignados a cada categoría de la PPT. 2. N°/proporción de pacientes con EP en cada categoría.

Tabla 2 Principales características de los trabajos que estudian una o varias escalas comparándolas con el empirismo. (Continuación)

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
<p>Carrier et al. 2006 (29)</p> <p>535 pacientes analizados</p>	<p>Estudio retrospectivo sin grupo control</p>	<p>Dos centros hospitalarios.</p> <p>Combinación de dos bases de datos de los dptos de Medicina Nuclear: pacientes consecutivos hospitalizados y ambulatorios con sospecha de EP.</p> <p>Criterios exclusión: menores de 18 años, incapacidad para dar consentimiento, expectativas de supervivencia inferiores a 3 meses, ventilación mecánica, EP crónico ó EP en los 3 últimos meses, tratamiento anticoagulante, filtro de vena cava inferior.</p> <p>Los pacientes analizados: 34,6% varones, edad mediana 52 años (rango 17-96).</p>	<p>Estimación empírica prospectiva de la PPT asignando la PPT a una de las siguientes categorías: menor del 20%, 20-80% y mayor del 80%.</p> <p>Estimación de la PPT a través de 3 modelos estructurados: la escala de Ginebra (17), la escala de Wells (16) y el modelo de Rodger (19). La escala de Wells se aplicó en dos formas: PPT en 3 categorías y PPT en dos categorías.</p> <p><i>Para aumentar la comparabilidad entre las diferentes formas de estimar la PPT, ésta fue categorizada en s en PPT baja y PPT no baja (PPT moderada y PPT alta).</i></p> <p>Se determinó el DD y realizó GVP a todos los pacientes.</p>	<p>Estándar de referencia: algoritmo diagnóstico que incluye GVP, UC, angiografía, angioTAC y seguimiento de 3 meses. Además incluye la PPT (escala Wells dicotomizada) y el valor del DD como elementos de decisión.</p> <p>EP presente: PPT alta/intermedia y GVP de alta probabilidad, angiografía positiva, TAC helicoidal positivo, UC positiva.</p> <p>EP descartado: DD negativo y baja PPT, GVP normal o casi normal, GVP de baja probabilidad y baja PPT, GVP de baja probabilidad con UC negativa, ausencia de ETE en el seguimiento.</p>	<p>Para cada una de 4 formas de estimar la PPT se ha determinado:</p> <ol style="list-style-type: none"> Nº/proporción de pacientes asignados a cada categoría de la PPT. Nº/proporción de pacientes con EP en cada categoría. <p>Para el modelo de Rodger se presentan sensibilidad, especificidad, valor predictivo negativo, razones de probabilidad positiva y negativa.</p>
<p>Penaloza et al. 2007 (30)</p> <p>291 pacientes incluidos</p> <p>185 pacientes analizados</p>	<p>Estudio prospectivo sin grupo control</p>	<p>Hospital docente.</p> <p>Pacientes consecutivos mayores de 18 años con sospecha de EP: ambulatorios.</p> <p>Criterios exclusión: terapia anticoagulante más de 48 horas, inaccesibilidad a GVP en primeras 24 horas, insuficiencia renal, rehusé ó incapacidad para consentimiento, embarazo, imposibilidad para seguimiento, expectativa de vida inferior a 3 meses, alergia a contraste iodado, contraindicación de anticoagulación, indicación trombolisis.</p> <p>Los pacientes analizados: 41,1% varones, edad mediana 56 años (rango= 18-94).</p>	<p>Estimación de la PPT a través de un modelo estructurado, la escala de Wells (16), por médicos en formación.</p> <p>Estimación de la PPT a través de la escala de Wells, aplicada por los mismos médicos, que en caso de duda consultan con sus supervisores (médicos experimentados), cuyo criterio prevalece si hay discordancia.</p> <p>La escala se utilizó en sus dos formas: clasificación de PPT en 3 categorías (prospectivamente) y PPT en 2 categorías (retrospectivamente).</p>	<p>Estándar de referencia: algoritmo diagnóstico que incluye DD, GVP, UC y TAC helicoidal. Además incluye la PPT como elementos de decisión.</p> <p>Todos los pacientes fueron seguidos durante 3 meses.</p> <p>EP presente: GVP de alta probabilidad, UC con TVP proximal, TAC helicoidal positivo, ocurrencia de un ETE durante el seguimiento.</p> <p>EP descartado: DD negativo, GVP normal, UC negativa y PPT baja, TAC helicoidal negativa, ausencia de ETE en seguimiento.</p>	<p>Para cada una de las 4 formas de estimar la PPT se ha determinado:</p> <ol style="list-style-type: none"> Nº/proporción de pacientes asignados a cada categoría de la PPT. Nº/proporción de pacientes con EP en cada categoría.

Tabla 3 Principales características de los trabajos que estudian varias escalas comparándolas entre sí.

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
<p>Chagnon et al. 2002 (25)</p> <p>390 pacientes incluidos</p> <p>277 pacientes analizados</p>	<p>Estudio prospectivo sin grupo control</p>	<p>Tres hospitales docentes.</p> <p>Pacientes consecutivos con sospecha de EP: departamento de urgencias.</p> <p>Criterios exclusión: datos insuficientes para calcular las escalas, protocolo diagnóstico incompleto.</p> <p>Los pacientes analizados: 44% varones, edad mediana 63 años (rango: 19-96).</p>	<p>Estimación de la PPT a través de dos modelos estructurados: la escala de Ginebra (17), de forma prospectiva, y la escala de Wells (16), de manera retrospectiva.</p> <p>Estimación de la PPT a través de la escala de Ginebra con superposición del juicio clínico en caso de discrepancia.</p> <p>Realizada por médicos residentes rotantes en el dpto. de urgencias.</p> <p><i>Se aplicó la escala de Ginebra con solapamiento del juicio clínico en caso de discrepancia, como es habitual en la utilización de este modelo.</i></p>	<p>Estándar de referencia: algoritmo diagnóstico que incluye DD, GVP, UC y angiografía. Además incluye la PPT (escala Ginebra) como elemento de decisión y seguimiento durante 3 meses.</p> <p>Todos los pacientes entran en el algoritmo.</p> <p>EP presente: TVP proximal en UC, TAC helicoidal positivo, angiografía anormal, GVP alta probabilidad, presentación de ETE en seguimiento.</p> <p>EP descartado: DD (ELISA) negativo, TAC helicoidal negativo en pacientes con PPT baja o intermedia, GVP normal o casi normal, angiografía negativa, ausencia de ETE en seguimiento.</p>	<p>Para cada una de 3 formas de estimar la PPT se ha determinado:</p> <ul style="list-style-type: none"> •Nº/proporción de pacientes asignados a cada categoría de la PPT. •Nº/proporción de pacientes con EP en cada categoría. <p>Área bajo la curva ROC incluyendo todos los posibles valores de la escala y categorizando la PPT en tres valores.</p>
<p>Moore et al. 2004 (31)</p> <p>295 pacientes incluidos</p> <p>295 pacientes analizados</p>	<p>Estudio retrospectivo sin grupo control</p>	<p>Centro de servicios sanitarios para militares.</p> <p>Análisis retrospectivo de datos de pacientes con sospecha de EP remitidos para GVP ó angioTAC: hospitalizados y ambulatorios.</p> <p>Criterios exclusión: menores de 18 años, sospecha de TVP aislada.</p> <p>Los pacientes analizados: 52% varones, edad media 60 años (DE= 18).</p>	<p>Estimación de la PPT a través de dos modelos estructurados: la escala de Ginebra (17) y la escala de Wells (16).</p> <p><i>Para la escala de Ginebra se realizaron dos grupos: los pacientes para los que estaba claro que la gasometría arterial se había realizado respirando aire ambiental se asignaron al grupo de "escala de Ginebra pura"; los pacientes de los que no había referencia explícita sobre si respiraban aire ambiental o suplementado con oxígeno en el momento de la gasometría se asignaron al grupo de "escala de Ginebra supuesta".</i></p>	<p>Estándar de referencia: cualquiera de las pruebas complementarias realizadas para el diagnóstico.</p> <p>EP presente: angioTAC positivo, angiografía positiva, UC MMII positiva en contexto de síntomas sugestivos de EP, GVP de alta probabilidad en el contexto de una PPT intermedia o alta.</p>	<p>Para cada una de 3 formas de estimar la PPT se ha determinado:</p> <ul style="list-style-type: none"> •Nº/proporción de pacientes asignados a cada categoría de la PPT. •Nº/proporción de pacientes con EP en cada categoría. <p>Para la escala de Wells se ha determinado la relación entre la PPT y la incidencia de EP en cada categoría.</p>

Tabla 3 Principales características de los trabajos que estudian varias escalas comparándolas entre sí. (Continuación)

Autor/año	Diseño	Población	Intervención	Estándar de diagnóstico	Medida de resultados clave
<p>Miniati et al. 2005 (32) 248 pacientes incluidos 215 pacientes analizados</p>	<p>Estudio prospectivo sin grupo control</p>	<p>Institución italiana.</p> <p>Pacientes consecutivos con sospecha de EP: hospitalizados y ambulatorios.</p> <p>Criterios exclusión: incapacidad para obtener el consentimiento informado, contraindicación de angiografía, tratamiento anticoagulante durante más de 3 días.</p> <p>Los pacientes analizados: 45% varones, edad mediana 70 años (rango: 17-90), 22% pacientes ambulatorios.</p>	<p>Estimación de la PPT a través de tres modelos estructurados: la escala de Wells (16) y el modelo de Pisa (9), que son aplicados de forma prospectiva, y la escala de Ginebra (17), que se calcula retrospectivamente.</p> <p>Realizada por neumólogos.</p> <p><i>En el modelo de Pisa se ha categorizado la PPT en 3 valores (baja <= 10%, intermedia >10 % - <= 90% y alta > 90%), aunque originalmente son cuatro las categorías.</i></p>	<p>Estándar de referencia: angiografía pulmonar. A todos los pacientes se les realizó gammagrafía pulmonar y fueron seguidos durante 1 año. EP presente: angiografía anormal, presentación de ETE en seguimiento. EP descartado: angiografía normal, ausencia de ETE en seguimiento.</p>	<p>Para cada una de 3 formas de estimar la PPT se ha determinado:</p> <ul style="list-style-type: none"> •Nº/proporción de pacientes asignados a cada categoría de la PPT. •Nº/proporción de pacientes con EP en cada categoría. •Área bajo curva ROC <p>Adicionalmente, se ha relacionado la extensión de la obstrucción vascular en la GVP con las 3 categorías de la PPT.</p>
<p>Ollenberger et al. 2006 (33) 1487 pacientes incluidos 1359 pacientes analizados</p>	<p>Estudio retros-pectivo sin grupo control</p>	<p>6 hospitales terciarios .</p> <p>Datos del estudio PIOPED (6,7): pacientes de 18 años ó más con sospecha de EP sin contraindicación para angiografía, hipersensibilidad al contraste, inestabilidad hemodinámica, hipertensión pulmonar conocida ó embarazadas.</p> <p>Criterios exclusión: pacientes con diagnóstico incierto de EP.</p> <p>Los pacientes analizados: 45% varones, edad media 55 años (DE= 18), 32% pacientes ambulatorios.</p>	<p>Estimación de la PPT a través de dos modelos estructurados: la escala de Ginebra (17) y la escala de Wells (16).</p> <p><i>Los pacientes fueron estratificados según su ubicación en el momento de la entrada al estudio PIOPED. El grupo de pacientes ambulatorios quedó constituido por los pacientes ambulatorios y de urgencias y el grupo de pacientes hospitalizados quedó constituido por los pacientes del ámbito quirúrgico y los de las unidades de cuidados médicos / coronarios y las unidades de cuidados intensivos.</i></p>	<p>Estándar de referencia: angiografía pulmonar. Todos los pacientes fueron seguidos durante 1 año. EP presente: angiografía pulmonar anormal.</p> <p>EP descartado: angiografía pulmonar negativa, GVP normal o de baja o muy baja probabilidad, ausencia de ETE en seguimiento.</p>	<p>Para cada uno de los 2 modelos y las diferentes ubicaciones de los pacientes se ha determinado:</p> <ul style="list-style-type: none"> •Nº/proporción de pacientes asignados a cada categoría de la PPT. •Nº/proporción de pacientes con EP en cada categoría. •Área bajo la curva ROC .

Valoración crítica y resultados de los estudios seleccionados

Para cada grupo de estudios, se presenta la valoración de la calidad de los trabajos con la herramienta QUADAS y seguidamente los resultados clave sobre el rendimiento diagnóstico.

> Estudios que analizan una escala ó modelo predictivo sin otra prueba diagnóstica de comparación

Los resultados de la evaluación de la calidad de estos estudios se presentan en la tabla 4.

Con respecto a los dos estudios que analizan el rendimiento diagnóstico de una escala (24, 26), su calidad, vista globalmente, podría considerarse aceptable. La validez externa, explorada a través de los ítems 1, 2 y 12, ha sido valorada como adecuada. En cuanto a la validez interna, es en los ítems 5, 6 y 7, que exploran la adecuación y uso del estándar de referencia, en los que se ha visto menor cumplimiento.

En cuanto al otro estudio de este grupo, el trabajo de Miniati (27), es necesario aclarar algunos aspectos para entender la evaluación. Se trata de un estudio realizado sobre una muestra consecutiva de pacientes, dirigido a estudiar el rendimiento de un algoritmo “no invasivo” para el diagnóstico del embolismo pulmonar (EP). Sin embargo, para esta revisión, sólo se ha tomado el resultado en el que se relacionan las diferentes categorías de probabilidad pretest (PPT) con el grado de severidad de la enfermedad. En este resultado sólo han sido considerados aquellos pacientes con embolismo pulmonar confirmado, que constituirían la población de referencia, lo que compromete la validez externa de la parte del estudio que interesa a esta revisión. Ello justifica que, aunque la puntuación global obtenida haya sido alta, se considere muy mermada la aplicabilidad del estudio.

Los dos estudios que analizan el rendimiento diagnóstico de la escala de Wells, sin establecer comparación con ningún otro procedimiento para estimar la probabilidad pretest, lo hacen sobre poblaciones con diferentes frecuencias de EP. El trabajo de Wells (24) se realizó sobre una población ambulatoria (pacientes atendidos en los servicios de urgencias) con una frecuencia estimada de EP de 9,5%. En estas condiciones, algo más de la mitad de los pacientes fueron asignados al grupo de baja PPT, siendo la proporción de pacientes con EP confirmado en esta categoría del 1,3% (Intervalo de confianza (IC) al 95% de 0,5-2,7), lo que resulta en un alto valor predictivo negativo (98,7%). La proporción de pacientes diagnosticados de EP en las categorías de PPT moderada y alta fue de 16,2% y 40,6%, respectivamente. La escala de Wells aplicada a esta población con baja frecuencia de EP y comparando la categoría de baja PPT frente a las categorías de moderada/alta PPT, se caracteriza por una alta sensibilidad (91,9%) y baja especificidad (61,6%).

Tabla 4	Evaluación con la herramienta QUADAS de la calidad de los estudios que analizan una escala
---------	--

ITEM de QUADAS	Wells 2001(24)	Christopher Study Investigator s 2006 (26)	Miniati 2003(27)
1. Espectro de pacientes representativo	Sí	Sí	No ¹
2. Descripción clara de los criterios de selección	Sí	Sí	Sí
3. Estándar de referencia adecuado	Sí	Sí	Sí
4. Transcurre poco tiempo entre prueba y estándar	Sí	Sí	Sí
5. Estándar aplicado a toda la muestra	No ²	No ²	Sí ³
6. Los pacientes reciben el mismo estándar	No	No	Sí
7. Estándar independiente de la prueba índice	No ⁴	Sí	Sí
8. Adecuada descripción de la prueba índice	Sí	Sí	Sí
9. Adecuada descripción del estándar	Sí	Sí	Sí
10. Interpretación de resultado de prueba independiente de resultado de estándar	Sí	Sí	Sí
11. Interpretación de resultado de estándar independiente de resultado de prueba	No claro ⁵	Sí	No claro ⁶
12. Disponibilidad de los mismos datos en la práctica	Sí	Sí	Sí
13. Descripción de resultados intermedios	Sí	Sí	Sí
14. Descripción de las pérdidas y salidas del estudio	Sí	Sí	Sí
Resumen	10 "sí"	12 "sí"	12 "sí"

¹ El objetivo de este estudio que interesa a esta revisión se ha llevado a cabo sobre aquellos pacientes que estaban diagnosticados de embolismo pulmonar, por lo que no constituye una muestra representativa de la población a la que será aplicada la prueba.

² A los pacientes con probabilidad pretest determinada como baja ó "poco probable" y dímero D negativo se les descarta el embolismo pulmonar sin realizarles más pruebas complementarias.

³ A todos los casos se les determina la probabilidad pretest (PPT) y se les realiza gammagrafía pulmonar (protocolo no invasivo que se está evaluando), aunque sólo en determinados casos se realiza la angiografía pulmonar. De este trabajo interesa un objetivo concreto, el que estudia la relación entre la PPT determinada y la severidad de la enfermedad en enfermos con diagnóstico confirmado de EP por angiografía, por lo que se puede considerar que todos han sido examinados con la prueba estándar.

⁴ La probabilidad pretest estimada a través de escala está incluida en uno o varios pasos del algoritmo de diagnóstico.

⁵ Deja claro que la gammagrafía de ventilación perfusión fue interpretada sin conocer el resultado de la probabilidad pretest, pero no lo deja claro en el caso de ultrasonografía y angiografía.

⁶ El personal entrenado que interpretó los resultados de la gammagrafía estaba cegado a los datos clínicos, pero no hace referencia a si la interpretación de la angiografía se hizo sin los datos de la probabilidad pretest, que sí fueron utilizados para realizar la indicación de esta prueba.

La frecuencia de EP en el trabajo de Christopher Study Investigators (26) es mayor (20,39%), al tratarse de una población en la que se mezclan pacientes ambulatorios y hospitalizados. En este estudio, la escala de Wells es aplicada de forma que la PPT es dicotomizada. Aproximadamente dos tercios de la población fueron asignados a la categoría de “EP improbable”, registrándose entre éstos un 12,1% de EP (IC 95% 10,7-13,5). La proporción de pacientes con EP entre los que habían sido categorizados como “EP probable” fue del 37,1% (IC 95% 34,2-40,0). Al categorizarse la PPT en dos posibles valores (estableciendo el punto de corte en 4 puntos), la escala se presenta como más específica que sensible, si bien, en esta población, ninguno de los dos parámetros alcanza el 80%.

Los resultados del trabajo de Miniati (27) que interesan a esta revisión implican sólo a pacientes con diagnóstico confirmado de EP. De este modo, aunque demuestran que existe una relación estadísticamente significativa entre la graduación de la PPT, estimada a través del modelo de Pisa, y el grado de severidad de la enfermedad (más baja PPT, menor severidad y más alta PPT mayor grado de severidad), los resultados no son extrapolables a ningún contexto diferente al de este estudio.

En la tabla 5 se encuentran más detallados los resultados de estos tres estudios.

Tabla 5. Resultados relacionados con el uso de escalas o modelos clínicos de los estudios incluidos en la revisión sistemática. Estudios sobre una escala o modelo.

Autor/año, población estudiada y frecuencia EP	Modelo para la estimación de la PPT	Categoría de PPT	Pacientes en cada categoría de PPT			Pacientes con EP en cada categoría de PPT		
			Nº	%	IC 95%	Nº	%	IC 95%
Wells et al. 2001 (24) 930 pacientes 9,5% EP	E Wells	Baja (<2) ^b	527	57		7	1,3	0,5-2,7
		Intermedia ^b	339	36		55	16,2	12,5-20,6
		Alta (>6) ^b	64	7		24	40,6	28,7-53,7
C. S. I. ^a 2006 (26) 3306 pacientes 20,39% EP	E Wells	Improbable ^b (≤ 4)	2206	66,7		266	12,1	10,7-13,5
		Probable ^b	1100	33,3		408	37,1	34,2-40,0
Miniati et al. ^c 2003 (27) 390 pacientes 41% EP	Modelo de Pisa	Baja						
		Intermedia						
		Moderada-mente alta						
		Alta						

E= escala; EP= embolismo pulmonar; IC= intervalo de confianza; PTT= probabilidad pretest.

^aChristopher Study Investigators. ^bLa diferencia entre categorías en la frecuencia de EP es estadísticamente significativa ($p < 0,001$). ^cEn este trabajo se evalúa un algoritmo diagnóstico amplio del que forma parte el modelo de Pisa. La exactitud diagnóstica ha sido evaluada para el algoritmo completo, por lo que no pueden aportarse datos del rendimiento diagnóstico del modelo de Pisa.

> Estudios que analizan una escala ó modelo predictivo comparándolas con el empirismo

La calidad de los trabajos de este grupo es muy variable, oscilando la puntuación resumen (número de respuestas "Sí") entre el 6 obtenido por el estudio de Carrier (29) y el 13 del estudio de Sanson (23). En la tabla 6 se muestran los resultados de la valoración de los 4 estudios del grupo.

Las cuestiones sobre la validez externa han sido valoradas positivamente en los 4 estudios. La mayor variabilidad se encuentra en aquellos ítems que se refieren a la definición y utilización del estándar de referencia. Pese a que los cuatro estudios emplean algoritmos para el diagnóstico definitivo de EP, hay mucha variedad en cuanto a las pruebas consideradas, la inclusión del valor de la probabilidad pretest como elemento de decisión en el algoritmo, o la inclusión del seguimiento en el mismo. Por último, la escasa puntuación del trabajo de Carrier (29) puede atribuirse, no sólo a los aspectos relacionados con el estándar de referencia, sino también a su diseño retrospectivo y a la falta de datos sobre la metodología de las pruebas índice y estándar, y sobre el cegamiento a la hora de interpretar los resultados de una y otra.

En el estudio de mejor calidad, el de Sanson (23), se evalúan el algoritmo de Wells y la escala de Wells frente al empirismo, considerando para este último tres posibles valores de la PPT. Se trata de un estudio realizado sobre una población mezclada de pacientes hospitalizados y ambulatorios, con alta frecuencia de EP (31%). La proporción de pacientes asignados a la categoría de baja probabilidad de EP es diferente según cual de los tres métodos de estimación de la PPT se haya usado, correspondiendo a ambas escalas las proporciones más altas (25% en el caso del algoritmo y 36% para la escala) y al empirismo la más baja (14%). La proporción de pacientes con EP en la categoría de baja PPT (falsos negativos) es inferior cuando dicha probabilidad es estimada mediante el empirismo (19%), que cuando se utilizan las escalas (28%), resultando superior la sensibilidad de éste. Por otro lado, las escalas presentan una mayor especificidad que el cálculo mediante el empirismo, si bien los valores de este parámetro no son altos para el algoritmo ni para la escala de Wells (29,1% y 36,3%, respectivamente).

En el estudio de Kline (28), en el que la frecuencia de EP es baja (4,7%), al tratarse de pacientes ambulatorios de cualquier grupo de edad, se vuelven a comparar la escala de Wells (con tres categorías) y el método empírico, si bien en esta ocasión, los niveles de probabilidad del empirismo son categorizados de forma diferente. La proporción de pacientes asignados a cada categoría de PPT es similar para ambos métodos de estimación, siendo también similar la proporción de pacientes diagnosticados de EP en cada categoría, y resultando, por tanto, similares parámetros de rendimiento diagnóstico. Tanto la especificidad, situada en torno al 73-75%, como la sensibilidad, en torno al 53-59%, podrían considerarse bajas, pero el valor predictivo negativo (VPN) determinado para la categoría de baja probabilidad resulta muy alto para ambos métodos, estimándose en un 97%, aproximadamente.

Tabla 6 Evaluación con la herramienta QUADAS de la calidad de los estudios que analizan escalas comparándolas con el empirismo

ITEM de QUADAS	Sanson 2000(23)	Kline 2006(28)	Carrier 2006(29)	Penalosa 2007(30)
1. Espectro de pacientes representativo	Sí	No claro ¹	Sí	Sí
2. Descripción clara de los criterios de selección	Sí	Sí	Sí	Sí
3. Estándar de referencia adecuado	Sí	No ²	Sí	Sí
4. Transcurre poco tiempo entre prueba y estándar	Sí	No claro ³	No aplicable ⁴	Sí
5. Estándar aplicado a toda la muestra	Sí ⁵	Sí	Sí	No ⁶
6. Los pacientes reciben el mismo estándar	Sí ⁵	Sí	No ⁷	Sí ⁵
7. Estándar independiente de la prueba índice	Sí	Sí	No ⁸	No ⁸
8. Adecuada descripción de la prueba índice	Sí	Sí	No claro ⁹	Sí
9. Adecuada descripción del estándar	Sí	Sí	No claro	No claro
10. Interpretación de resultado de prueba independiente de resultado de estándar	Sí	Sí	No claro ¹⁰	Sí
11. Interpretación de resultado de estándar independiente de resultado de prueba	No claro	Sí	No claro	No claro ¹¹
12. Disponibilidad de los mismos datos en la práctica	Sí	Sí	Sí	Sí
13. Descripción de resultados intermedios	Sí	Sí	Sí	Sí
14. Descripción de las pérdidas y salidas del estudio	Sí	Sí	No	Sí
Resumen	13 "sí"	11 "sí"	6 "sí"	10 "sí"

¹ Incluye a todos los grupos de edad, sin hacer referencia a la cantidad de población no adulta incluida. No obstante se ha decidido la inclusión de este trabajo en nuestro estudio basándonos en el hecho de que las manifestaciones de la enfermedad tromboembólica son muy raras en la infancia/adolescencia, de modo que la población estará mayoritariamente constituida por personas adultas.

² Los autores hacen explícito que el estándar de referencia para todos los pacientes es el seguimiento durante 90 días tras la presentación del evento, pero algunos pacientes son examinados mediante pruebas de imagen consideradas de referencia para el diagnóstico del EP, dependiendo de la probabilidad pretest y el resultado del dímero D.

³ El período de tiempo fijado para el seguimiento (criterio estándar para todos los pacientes en este estudio) es el adecuado para la confirmación o el descarte de aquellos casos que no fueron inicialmente considerados como embolismo pulmonar, ya que es lo suficientemente prolongado como para que la enfermedad progrese, y en ese caso habría que responder "no". Sin embargo, a determinados pacientes se les somete a pruebas de imagen para confirmar el diagnóstico, y para éstos el seguimiento tiene un objetivo distinto. Los autores no especifican la demora existente entre la valoración clínica y la realización de las pruebas complementarias, por lo que la respuesta en este ítem sería "no claro".

⁴ Es un estudio retrospectivo. La comprobación del estado de enfermedad mediante pruebas estándar fue anterior al cálculo de la probabilidad pretest mediante escalas.

⁵ Todos los pacientes son incluidos en el algoritmo diagnóstico considerado estándar de referencia, aunque no a todos los pacientes se les realicen las mismas pruebas.

⁶ A los que resultaron tener dímero D negativo no se les realizaron más pruebas diagnósticas, salvo el seguimiento.

⁷ El resultado de la probabilidad pretest y/o del dímero D determinan la entrada en el algoritmo diagnóstico.

⁸ La probabilidad pretest estimada a través de escala está incluida en uno o varios pasos del algoritmo de diagnóstico.

⁹ No que queda claro qué escalas son calculadas de manera prospectiva y cuáles de manera retrospectiva.

¹⁰ Los autores no hacen referencia a que los investigadores que calcularon las escalas de forma retrospectiva estuvieran cegados a los resultados de las pruebas de imagen.

¹¹ No dice explícitamente que en la interpretación de las pruebas complementarias estuvieran cegados a los resultados clínicos, aunque la probabilidad pretest es un factor que determina la indicación de determinadas pruebas.

Muy similar, 96%, resulta el VPN estimado para la regla de Charlotte, el tercer método de cálculo de la PPT con el que se comparan los dos métodos anteriores en este mismo trabajo (28). Sin embargo, la especificidad de este modelo es mucho mayor que la

resultante para el empirismo y la escala de Wells (91%, aproximadamente), al tratarse de un modelo con sólo dos categorías, que asigna a los pacientes de forma mayoritaria a la categoría negativa (90,2%), en un ámbito de baja frecuencia de EP.

El estudio de Carrier (29) resulta difícil de valorar ya que se comparan 5 métodos de determinación de la PPT y cada determinación se ha realizado sobre una población diferente (coinciden para las dos formas de aplicar la escala de Wells), si bien las frecuencias de EP son similares entre ellas, en un rango que oscila entre el 17,9% (muestra sobre la que se utilizó la escala de Ginebra) y el 19,8% (para el modelo de Rodger). Si se comparan los diferentes métodos en términos de sensibilidad y especificidad los más similares resultan ser el empirismo y la escala de Wells con punto de corte en la puntuación 4, con sensibilidad superior al 80% y especificidad en torno al 38-40%. Los modelos más sensibles resultan ser la escala de Wells con punto de corte en 2 y el modelo de Rodger, con sensibilidad superior al 90%, si bien el modelo de Rodger se muestra algo más específico que el de Wells (28,8% del primero, frente a 19,3% del segundo). La escala de Ginebra es el único método que se muestra más específico que sensible, hallazgo que no se ha constatado en otros estudios en los que se ha empleado dicha escala (25, 31-33). La escasa calidad metodológica del estudio, evidenciada al ser evaluado con la herramienta QUADAS, compromete la validez de estos resultados.

El trabajo de Penalzoa (30) plantea el estudio de la escala de Wells en sus dos formas de categorizar la PPT desde un punto de vista diferente, pues se compara el rendimiento de la escala cuando es aplicada por médicos sin experiencia, con la utilización de la misma escala aplicada por los mismos médicos, pero añadiendo la valoración crítica de un médico experimentado (por lo que ha sido equiparado al empirismo). En ambas situaciones, la escala con tres categorías se muestra más sensible que específica, situándose la sensibilidad por encima del 90% (en ambos casos el IC 95% contiene al 1), y la especificidad en torno al 63-65%. Sin embargo, cuando se aplica la escala de Wells con dos categorías, ésta resulta ser más específica (valores superiores al 80%) que sensible.

Tanto si se divide la PPT estimada en dos categorías, como en tres, la valoración adicional por parte del clínico experimentado, con respecto a la utilización de la escala sólo por clínicos no experimentados, supone un incremento en los dos parámetros de rendimiento referidos, si bien la diferencia no es estadísticamente significativa.

Los principales resultados de estos cuatro estudios se presentan en la tabla 7 que se muestra a continuación.

Tabla 7 Resultados relacionados con el uso de escalas o modelos de los estudios incluidos en la revisión sistemática. Estudios que evalúan escalas frente a empirismo.

or/año, población estudiada y % EP	Estimación PPT (n)	Categoría de PPT	Pacientes en cada categoría de PPT			Pacientes con EP en cada categoría de PPT			Parámetros estimados sobre la exactitud diagnóstica Otros resultados relacionados con el uso de escalas
			Nº	%	IC 95%	Nº	%	IC 95%	
Sansón et al. 2000 (23) 517 pacientes 31% EP	Empirismo (413)	< 20%	58	14		11	19	9-29	LR= 0,53 (IC 95% 0,29-0,99)
		20-80%	278	67		80	29	23-34	LR= 0,92 (IC 95% 0,79-1,07)
		>80%	77	19		35	46	34-57	LR= 1,90 (IC 95% 1,28-2,82)
	E Wells (414)	Baja	147	36		41	28	21-35	LR= 0,93 (IC 95% 0,69-1,24)
		Moderada	259	63		78	30	25-36	LR= 1,03 (IC 95% 0,88-1,21)
		Alta	8	2		3	38	4-71	LR= 1,44 (IC 95% 0,35-5,92)
	Algoritmo de Wells (237)	Baja	60	25		17	28	17-40	LR= 0,66 (IC 95% 0,40-1,08)
		Moderada	138	58		54	39	31-47	LR= 1,07 (IC 95% 0,86-1,33)
		Alta	39	16		18	46	31-62	LR= 1,43 (IC 95% 0,80-2,53)
Kline et al. 2006. (28) 2302 pacientes 4,7% EP	Empirismo (2302)	< 15%	1650	71,7		44	2,7	1,9-3,6	A partir de los datos originales se han estimado SE= 0,593 (IC 95% 0,500-0,685) ES= 0,732 (IC 95% 0,713-0,751)
		15-40%	605	26,3		46	7,6	5,5-9,7	
		> 40%	47	2,0		18	38,3	24,4-52,2	
	E Wells (2302)	Baja (≤ 2)	1704	74,0		50	2,9	2,2-3,9	A partir de los datos originales se han estimado SE= 0,537 (IC 95% 0,443-0,631) ES= 0,754 (IC 95% 0,736-0,772)
		Moderada	559	24,3		48	8,6	6,3-10,9	
		Alta	39	1,7		10	25,6	11,9-39,3	
R Charlotte (2302)	Negativa	2077	90,2		81	3,9	3,1-4,8	A partir de los datos originales se han estimado SE= 0,250 (IC 95% 0,168-0,332) ES= 0,910 (IC 95% 0,898-0,922)	
	Positiva	225	9,8		27	12,0	7,8-16,2		

Tabla 7 Resultados relacionados con las escalas o modelos de los estudios incluidos en la revisión sistemática. Estudios que evalúan escalas frente a empirismo (continuación).

Autor/año, población estudiada y % EP	Estimación PPT (n)	Categoría de PPT	Pacientes en cada categoría de PPT			Pacientes con EP en cada categoría de PPT			Parámetros estimados sobre la exactitud diagnóstica Otros resultados relacionados con el uso de escalas
			Nº	%	IC 95%	Nº	%	IC 95%	
Carrier et al. 2006 (29) 535 pacientes 18% EP	Empirismo (432)	Baja (<20%)	116	23,9		11	7,5 ^b	2,2-11,9 ^b	A partir de los datos originales se han estimado SE= 0,859 (IC 95% 0,782-0,936) ES= 0,381 (IC 95% 0,331-0,432)
		No baja ^a	286	66,2		67	23,4 ^b	18,5-28,3 ^b	
	E Ginebra (280)	Baja	104	60,2		21	10,9 ^b	6,5-15,2 ^b	A partir de los datos originales se han estimado SE= 0,580 (IC 95% 0,443-0,717) ES= 0,752 (IC 95% 0,696-0,808)
		No baja ^a	86	30,7		29	33,7 ^b	23,7-43,7 ^b	
	E Wells (413)	Baja (<2)	69	16,7		1	5,8 ^b	0,2-11,2 ^b	A partir de los datos originales se han estimado SE= 0,947 (IC 95% 0,897-0,998) ES= 0,193 (IC 95% 0,151-0,235)
No baja ^a		344	83,3		72	20,9 ^b	16,6-25,2 ^b		
E Wells (413)	Baja (<4)	150	26,2		12	9,7 ^b	4,2-12,2 ^b	A partir de los datos originales se han estimado SE= 0,829 (IC 95% 0,744-0,914) ES= 0,407 (IC 95% 0,354-0,459)	
M Rodger (399)	Baja	95	23,8		2	2,2 ^b	0,1-6,7 ^b	SE= 96,20 (IC 95% 89-99); ES= 28,8 (IC 95% 24-34); VPN= 96,8 (IC 95% 91-99); LR (+)= 1,54 (IC 95% 1,3-1,9); LR (-)= 0,31 (IC 95% 0,1-0,9); VN= 23,1 (IC 95% 19-28)	
Penaloza et al. 2007 (30) 185 pacientes 18,4% EP	E Wells ^c (185)	Baja	99	53,5		2	2	1,0	A partir de los datos originales se han estimado SE= 0,912 (IC 95% 0,816-1,007) ES= 0,634 (IC 95% 0,559-0,713)
		Moderada	80	43,2		25	21	22,12	
		Alta	6	3,2		6	100	61-100	
	E Wells + empirismo ^d (185)	Baja	104	51,6		2	2	1,7	A partir de los datos originales se han estimado SE= 0,941 (IC 95% 0,862-1,020) ES= 0,656 (IC 95% 0,580-0,731)
Moderada		77	41,6		25	22	22,11		
Alta		7	3,8		7	100	65-100		
E Wells ^e (185)	Improbable	112	76,8		12	9	5-15	A partir de los datos originales se han estimado SE= 0,618 (IC 95% 0,454-0,781) ES= 0,854 (IC 95% 0,798-0,911)	
	Probable	43	23,2		21	49	35-63		
E Wells + empirismo ^d (185)	Improbable	111	77,9		10	7	4-12	A partir de los datos originales se han estimado SE= 0,706 (IC 95% 0,553-0,859) ES= 0,887 (IC 95% 0,837-0,938)	
	Probable	41	22,2		24	59	43-72		

E= escala; EP= embolismo pulmonar; ES=especificidad; IC= intervalo de confianza; LR= razón de probabilidad; M= modelo; PTT= probabilidad pretest; SE= sensibilidad; VN= verdaderos negativos; VPN= valor predictivo negativo; VPP= valor predictivo positivo; sombreados se presentan los datos calculados a partir de los datos originales.

^a Categoría PPT “no baja” es el resultado de sumar los resultados de las categorías moderada y alta. ^b Los porcentajes y correspondientes intervalos de confianza que se reseñan no son los que figuran el trabajo original de Carrier et al. Éstos han sido estimados de forma secundaria a partir de las cifras absolutas originales que aparecen en este trabajo. ^c Las tasas de embolismo pulmonar entre los 3 grupos de PPT son diferentes de forma estadísticamente significativa. ^d El número total de pacientes incluye a 135 pacientes valorados sólo por los médicos en formación y 50 pacientes valorados por éstos con el asesoramiento del supervisor ($p=0.98$) ^eLas tasas de EP entre los 2 grupos de PPT son diferentes de forma estadísticamente significativa ($p<0.0001$)

> Estudios que analizan varias escalas comparándolas entre sí

La calidad de los estudios, valorada globalmente a través de la puntuación obtenida, ha resultado adecuada en dos de los estudios (25), (32). Los otros dos trabajos fallan en un mayor número de preguntas de la herramienta QUADAS (31-33). En la tabla 8 se presentan los resultados de la valoración de estos trabajos.

Los ítems que hacen referencia a la validez externa han sido puntuados positivamente en dos de los cuatro estudios (25, 32). Sin embargo, en el trabajo de Moores (31) se analiza una población de pacientes que fueron remitidos para la realización de pruebas complementarias por sospecha de EP, por lo que es de suponer que en esta muestra estén sobrerrepresentadas las probabilidades medias-altas (que justifican la utilización de pruebas invasivas). En el estudio de Ollenberger (33) el espectro de pacientes no puede considerarse totalmente representativo de la población en la que se aplicará la prueba, pues fueron excluidos aquellos pacientes en los que el diagnóstico de embolismo pulmonar no pudo ser definitivamente confirmado ó descartado.

La validez interna está igualmente más comprometida en estos dos últimos trabajos, considerados retrospectivos (31, 33). Los problemas en la valoración de estos estudios se derivan fundamentalmente de la falta de una adecuada descripción del estándar de referencia, y la ausencia de información sobre si existió cegamiento del resultado de la prueba estándar a la hora de interpretar el resultado de las escalas.

En el trabajo de Chagnon (25), se comparan las escalas de Wells (3 categorías) y Ginebra. Para la escala de Ginebra, la PPT es estimada de dos formas distintas; una que asigna a los pacientes a una categoría de PPT según la puntuación obtenida tras aplicar la escala, y otra, que asigna también según la puntuación obtenida, pero teniendo en cuenta el criterio del médico que evalúa el caso, que es el criterio que prevalece en caso de existir desacuerdo (en otros trabajos es la forma habitual de utilizar la escala por el grupo de autores que la desarrolló y validó). Los tres métodos evaluados asignan a las categorías de PPT baja e intermedia una proporción semejante de pacientes; sin embargo, la escala de Ginebra con solapamiento del juicio clínico (ó empirismo) asigna a la categoría de PPT alta una mayor proporción de pacientes (10%) que las escalas de Wells y Ginebra, cuando ésta se aplica sin el concurso del empirismo (4% en ambos casos). La proporción de pacientes diagnosticados de EP es similar para los tres modelos en la categoría de PPT moderada, pero difiere en las categorías de PPT baja y alta, sobre todo en ésta última. La proporción de falsos negativos es inferior para la escala de Ginebra con participación del empirismo (5%) que para las escalas de Ginebra sin empirismo y la escala de Wells (13% y 12%, respectivamente); en cambio, la proporción de verdaderos positivos (pacientes diagnosticados de EP en las categorías de moderada y alta PPT), es superior para la escala de Wells.

Resumiendo, podría decirse que en este estudio la escala más sensible es la de Ginebra con solapamiento del juicio clínico (88,7%), seguida de la de Wells (73,2%) y la escala de Ginebra sin concurso del empirismo (71,8%). Con respecto a la especificidad, la escala con mejor resultado es la de Wells (69,4%), seguida de la escala de Ginebra con

solapamiento del juicio clínico (67,0%), y la escala de Ginebra (64,1%). Las diferencias entre las escalas de Wells y Ginebra sin solapamiento del juicio clínico, no son significativas desde un punto de vista estadístico. A la hora de interpretar estos resultados es necesario considerar que se trata de un estudio realizado sobre una población de pacientes ambulatorios con alta frecuencia de EP (26%), ámbito muy similar a aquel en el que se desarrolló y validó la escala de Ginebra (cohorte de pacientes con sospecha de EP atendidos en el departamento de urgencias).

El estudio de Moores (31) se ha desarrollado sobre una población de pacientes ambulatorios y hospitalizados con una frecuencia de EP del 30% (es preciso tener en cuenta que se trataba de pacientes derivados para la realización de pruebas complementarias invasivas). En éste se comparan las escalas de Wells y Ginebra, que fueron calculadas de manera retrospectiva. Este cálculo retrospectivo ha dado pie a hacer dos estimaciones diferentes al aplicar la escala de Ginebra, según se tuviera constancia de que la gasometría arterial se había realizado en las condiciones ideales (escala de Ginebra “pura”) o éste hecho fuera asumido, en ausencia de información que indicase lo contrario (escala de Ginebra “asumida”). Aunque la proporción de pacientes clasificados en cada categoría de la PPT es similar para los tres métodos, la proporción de pacientes diagnosticados de EP es diferente para cada uno de ellos: en la categoría de baja probabilidad, la proporción de falsos negativos es superior para la escala de Ginebra “asumida” (38,1%), que para la escala de Ginebra “pura” (25,9%) y la escala de Wells (15,3%). Esta última, es la que se muestra más sensible (83%) y más específica (40%), alcanzando la categoría de baja PPT el valor predictivo negativo más alto (84,7%).

El trabajo de Miniati (32) se efectuó sobre una población pacientes ambulatorios y hospitalizados con sospecha de EP entre los que la frecuencia de enfermedad era 43,3% (muy alta, pero es preciso tener en cuenta que a todos los pacientes se realizó angiografía pulmonar). En él se comparan tres modelos: la escala de Ginebra, la escala de Wells y el modelo de Pisa, que en esta ocasión, para mejorar la comparabilidad con los otros dos métodos, presenta tres categorías de PPT, en lugar de las cuatro habituales. La proporción de pacientes clasificados en cada categoría de la PPT es diferente para cada escala. Si se compara el modelo de Pisa con respecto a los otros dos, éste clasifica a una mayor proporción de pacientes en el grupo de baja probabilidad, y en ellos, la proporción de pacientes con diagnóstico de EP es del 5%, lo que representa un alto valor predictivo negativo (95%). En la categoría de alta PPT, el modelo de Pisa ubica a un 26% de los pacientes (frente al 28% de la escala de Ginebra y 15% de la escala de Wells), entre los que la frecuencia de EP es del 98% (frente al 49 y 64% de las escalas de Ginebra y Wells, respectivamente). Comparándolas en términos de sensibilidad y especificidad, el método más sensible es el modelo de Pisa (95,7%, IC95% 91,6-99,5%), seguido de la escala de Wells (91,4%, IC 95% 85,7-97,1%) y la escala de Ginebra (86%, IC 95% 79,0-93,1%). El modelo de Pisa aparece también como el más específico (61,5%, IC 95% 52,8-70,1%), seguido por las escalas de Wells (45,9%, IC 95% 37,1-54,7%) y Ginebra (10,7%, IC 95% 5,2-16,1%).

En este estudio se presenta un resultado adicional, al relacionar las diferentes categorías de la PPT con el grado de severidad de la enfermedad. Sólo para el modelo de Pisa se establece una relación estadísticamente significativa entre ambas variables.

Tabla 8 Evaluación con la herramienta QUADAS de la calidad de los estudios que analizan escalas comparándolas entre sí

ITEM de QUADAS	Chagnon 2002 (25)	Moores 2004 (31)	Miniati 2005 (32)	Ollenberger 2006 (33)
1. Espectro de pacientes representativo	Sí	No claro ¹	Sí	No ²
2. Descripción clara de los criterios de selección	Sí	Sí	Sí	No claro ²
3. Estándar de referencia adecuado	Sí	Sí	Sí	Sí
4. Transcurre poco tiempo entre prueba y estándar	No claro ³	No aplicable ⁴	Sí	No aplicable ⁴
5. Estándar aplicado a toda la muestra	Sí ⁵	Sí ⁶	Sí	No ⁷
6. Los pacientes reciben el mismo estándar	Sí ⁵	No	Sí	No
7. Estándar independiente de la prueba índice	No ⁸	Sí	Sí	Sí
8. Adecuada descripción de la prueba índice	Sí	Sí	Sí	Sí
9. Adecuada descripción del estándar	Sí	No	Sí	Sí
10. Interpretación de resultado de prueba independiente de resultado de estándar	Sí	No claro	No claro	No claro
11. Interpretación de resultado de estándar independiente de resultado de prueba	No claro ⁹	Sí	Sí	Sí
12. Disponibilidad de los mismos datos en la práctica	Sí	Sí	Sí	Sí
13. Descripción de resultados intermedios	Sí	Sí	Sí	Sí ¹⁰
14. Descripción de las pérdidas y salidas del estudio	Sí	Sí	Sí	Sí
Resumen	11 "sí"	9 "sí"	13 "sí"	8 "sí"

¹ Estudio retrospectivo sobre una cohorte de pacientes con sospecha de EP que fueron remitidos para la realización de pruebas complementarias invasivas. Se plantea la duda de que en esta población se hayan incluido sujetos con probabilidades pretest más altas que las que se encontrarían en la muestra consecutiva de pacientes.

² Hace referencia a un artículo original y una revisión sobre el estudio PIOPED, cuya población se ha utilizado para este estudio, pero no explícita en la metodología los criterios de inclusión /exclusión del estudio original. Sí que hace referencia al criterio de exclusión en este estudio: pacientes a los que no se les pudo confirmar ó descartar el diagnóstico de embolismo pulmonar ("embolismo pulmonar incierto"), lo que compromete la representatividad de la muestra del estudio analizado.

³ Sólo especifica que la ultrasonografía se realiza en las primeras 24 horas desde la presentación, para el resto de las pruebas no aporta información.

⁴ Es un estudio retrospectivo. La comprobación del estado de enfermedad mediante pruebas estándar fue anterior al cálculo de la probabilidad pretest mediante escalas.

⁵ Todos los pacientes son incluidos en el algoritmo diagnóstico considerado estándar de referencia, aunque no a todos los pacientes se les realicen las mismas pruebas.

⁶ En este estudio retrospectivo no se hace referencia a la prueba complementaria que determinó el diagnóstico de embolismo pulmonar en cada caso, pero todos los pacientes fueron estudiados mediante alguna prueba de imagen

⁷ La población del estudio original fue randomizada en dos grupos; en uno se realizaba la angiografía pulmonar si el resultado de la GVP era interpretado como anormal, en el otro grupo la decisión de realizar la angiografía dependía del clínico que atendía el caso.

⁸ La probabilidad pretest estimada a través de escala está incluida en uno o varios pasos del algoritmo de diagnóstico.

⁹ Los técnicos que hicieron la determinación del dímero D estaban cegados a los resultados de la valoración clínica. Para la interpretación de las demás pruebas complementarias no hace referencia sobre si se utilizaron los resultados de la probabilidad pretest determinada, que sí se utilizaron para hacer la indicación.

¹⁰ En los trabajos originales a los que hace referencia se detallan los resultados intermedios.

Ollenberger y colaboradores comparan en su trabajo (33) las escalas de Wells y Ginebra cuando son aplicadas en diferentes ámbitos clínicos, para lo que consideran una población de pacientes ambulatorios y hospitalizados con sospecha de EP, entre los que la frecuencia de EP es alta (29%). Cuando ambas escalas son aplicadas a la población total, la escala de Ginebra se muestra más sensible que la de Wells (78,9%, IC 74,2-83,6% vs 73,2%, IC 95% 68,8-77,5%), pero menos específica (38,2%, IC 95% 34,6-41,8% vs 52,9%, IC 95% 49,8-56,1%). Si se considera como parámetro de rendimiento diagnóstico el área bajo la curva ROC, tanto para la población considerada globalmente, como para las subpoblaciones de pacientes ambulatorios y pacientes hospitalizados, la exactitud de la escala de Wells es mayor que la de la escala de Ginebra. Ambas escalas funcionan mejor cuando se utilizan en pacientes ambulatorios que en pacientes hospitalizados (diferencias estadísticamente significativas para ambos métodos). En el contexto del paciente hospitalizado, la exactitud es mayor entre los pacientes de unidades de cuidados médicos o intensivos, que entre los pacientes quirúrgicos.

En la tabla 9 se resumen los principales resultados de los cuatro estudios que comparan escalas o modelos entre sí.

Tabla 9 Resultados relacionados con las escalas o modelos clínicos de los estudios incluidos en la revisión sistemática. Estudios que comparan escalas entre sí.

Autor/año, población estudiada y % EP	Estimación PPT (n)	Categoría de PPT	Pacientes en cada categoría de PPT			Pacientes con EP en cada categoría de PPT			Parámetros estimados sobre la exactitud diagnóstica Otros resultados relacionados con el uso de escalas	
			Nº	%	IC 95%	Nº	%	IC 95%		
Chagnon et al. 2002 (25) 277 pacientes 26% EP	E Wells ^a (277)	Baja	162	58		19	12	7-17	ABC ROC E Wells continua: 0,78 (IC 95% 0,72-0,84) ABC ROC E Wells categorizada: 0,73 (IC 95% 0,66-0,80)	
		Moderada	104	38		42	40	31-50		
		Alta	11	4		10	91	74-108		
	E Ginebra ^a (277)	Baja	152	55		20	13	8-19		
		Moderada	113	41		43	38	29-47		ABC ROC E Ginebra continua: 0,74 (IC 95% 0,68-0,80) ABC ROC E Ginebra categorizada: 0,69 (IC 95% 0,62-0,76)
		Alta	12	4		8	67	40-93		
	E Ginebra + empirismo (277)	Baja	146	53		8	5	2-9		
		Moderada	103	37		41	40	30-49		ABC ROC E Ginebra con solapamiento del juicio clínico: 0,82 (IC 95% 0,76-0,87)
		Alta	28	10		22	79	63-94		

Tabla 9 Resultados relacionados con las escalas o modelos clínicos de los estudios incluidos en la revisión sistemática. Estudios que comparan escalas entre sí.

Autor/año, población estudiada y % EP	Estimación PPT (n)	Categoría de PPT	Pacientes en cada categoría de PPT			Pacientes con EP en cada categoría de PPT			Parámetros estimados sobre la exactitud diagnóstica Otros resultados relacionados con el uso de escalas
			Nº	%	IC 95%	Nº	%	IC 95%	
Moore et al. 2004 (31) 295 pacientes 30% EP	E Wells ^b (295)	Baja	98	33,2		15	15,3	9,5-23,7	A partir de los datos originales se han estimado SE= 0,830 (IC 95% 0,751-0,908) ES= 0,401 (IC 95% 0,334-0,468) Con respecto a la PPT baja, el riesgo de EP se multiplica por 2,95 en el grupo de PPT intermedia (IC 95% 1,56-5,59) y por 4,95 en el grupo de PPT alta (IC 95% 2,11-11,64).
		Moderada	161	54,6		56	34,8	27,9-42,4	
		Alta	36	12,2		17	47,2	32,0-63,0	
	E Ginebra "pura" ^c (79)	Baja	27	34,2		7	25,9	13,0-45,0	
		Moderada	42	53,2		10	23,8	13,0-38,0	
		Alta	10	12,6		6	60,0	31,0-83,0	
Ginebra "supuesta" ^c (159)	Baja	55	34,7		21	38,1	26,0-51,0	A partir de los datos originales se han estimado SE= 0,523 (IC 95% 0,375-0,670) ES= 0,296 (IC 95% 0,212-0,379)	
	Moderada	88	55,3		16	18,1	11,0-28,0		
	Alta	16	10,0		7	43,8	23,0-67,0		
Ollenberger et al. 2006 (33) 1359 pacientes 29% EP	E Wells ^e (1359)	Baja	615	45,3		107	17,4	14,4-20,4	ABC ROC total 0,68 (IC 95% 0,65-0,71) ABC ROC ambulatorios 0,76 (IC 95% 0,73-0,79) ABC ROC hospitalizados 0,65 (IC 95% 0,63-0,67) ^f
		Moderada	632	46,5		230	36,3	32,6-40,1	
		Alta	112	8,2		62	55,4	46,2-64,6	
	E Ginebra ^e (998)	Baja	332	33,3		61	18,4	14,2-22,5	
		Moderada	492	49,3		152	30,9	26,8-35,0	
Alta	174	17,4		76	43,7	36,3-51,0	ABC ROC hospitalizados 0,60 (IC 95% 0,57-0,63) ^g		

ABC= área bajo la curva; EP= embolismo pulmonar; ES=especificidad; IC= intervalo de confianza; LR= razón de probabilidad; M= modelo; PTT= probabilidad pretest; SE= sensibilidad; VN=

verdaderos negativos; VPN= valor predictivo negativo; VPP= valor predictivo positivo; sombreados se presentan los datos calculados a partir de los datos originales.

^a No hay diferencias entre el área bajo la curva ROC de la escala de Ginebra y la escala de Wells continuas ($p=0.21$); y tampoco entre ambas áreas cuando las escalas son categorizadas. Pero la escala de Ginebra solapada con juicio clínico se comporta mejor que la escala de Ginebra y la escala de Wells categorizadas de forma significativa.

^b La diferencia de incidencia de la enfermedad en los diferentes grupos es estadísticamente significativa ($p=0,001$). Si se consideran dos grupos de pacientes, los de PPT baja frente a PPT no baja (PPT intermedia y PPT alta), la diferencia de incidencia de la enfermedad entre ambos grupos continúa siendo estadísticamente significativa ($p=0,000$). El riesgo de presentar EP en el grupo de PPT no baja es 3,26 superior al de los pacientes del grupo de PPT baja (IC 95% 1,75-6,06).

^c La diferencia de incidencia de la enfermedad en los diferentes grupos no es estadísticamente significativa. Para cualquiera de las dos formas de calcular la escala de Ginebra, si se consideran los pacientes de PPT baja frente a los de PPT no baja (PPT intermedia y PPT alta), la diferencia de incidencia de la enfermedad entre ambos grupos tampoco alcanza la significación estadística.

^d Si se comparan los tres modelos por parejas, las diferencias entre las áreas bajo la curva ROC son todas estadísticamente significativas ($p < 0,0001$). Se encontró una relación altamente significativa ($p < 0,0001$) entre la probabilidad clínica predecida por el modelo de Pisa y la severidad del EP. El grado de obstrucción vascular fue mayor en los pacientes que estaban asignados a la categoría de alta PPT y fue menor en los pacientes asignados a la categoría de baja PPT. La relación entre la probabilidad clínica y la extensión de la embolización estaba en el límite de la significación para la escala de Wells ($p=0,005$) y fue no significativa para el modelo de Ginebra ($p=0,42$).

^e La diferencia existente entre ambas escalas en la proporción de pacientes clasificados como baja PPT y alta PPT es estadísticamente significativa ($p < 0,05$).

^f Ámbito quirúrgico ABC ROC $0,58 \pm 0,04$; cuidados médicos/coronarios ABC ROC $0,66 \pm 0,03$; unidades cuidados intensivos ABC ROC $0,69 \pm 0,05$. Diferencias estadísticamente significativas entre pacientes hospitalizados y ambulatorios en el ABC.

^g Ámbito quirúrgico ABC ROC $0,52 \pm 0,05$; cuidados médicos/coronarios ABC ROC $0,62 \pm 0,04$; unidades cuidados intensivos ABC ROC $0,56 \pm 0,08$. Diferencias estadísticamente significativas entre pacientes hospitalizados y ambulatorios en el ABC.

Síntesis de resultados sobre rendimiento diagnóstico

> 1. Escalas frente a empirismo

La utilidad del uso de escalas o modelos de predicción del riesgo de embolismo pulmonar (EP) frente al empirismo se ha estudiado en tres estudios de buena calidad (23, 28, 30) y uno de mala calidad (29). Los tres estudios de buena calidad incluyen la escala de Wells de tres categorías, uno de ellos añade la escala de Wells de dos categorías, otro el algoritmo de Wells y otro la regla de Charlotte. La frecuencia de EP es muy distinta en estos estudios (4,7%, 18% y 31%); sin embargo, tienen en común que las diferencias que encuentran entre el rendimiento diagnóstico del empirismo y de las escalas no son de gran magnitud. Aunque dos de estos estudios (Sanson (23), Kline (28)) encuentran que la escala de Wells resulta algo menos sensible que el empirismo, sólo en uno de ellos estas diferencias llegan a ser estadísticamente significativas y sólo para la sensibilidad (no para los valores predictivos). Este es el caso del estudio de Sanson (23), que halló una sensibilidad de la escala de Wells del 66% (IC 95%: 58-75%) y una sensibilidad del empirismo del 91% (IC 95%: 86-96%).

> 2. Rendimiento diagnóstico de las diversas escalas

2.1. Escala de Wells

La escala de Wells, sobre todo la modalidad de 3 categorías, ha resultado ser el modelo más estudiado. Por ello va a ser utilizada como punto de referencia en la síntesis de los resultados del estudio.

El rendimiento diagnóstico de la escala de Wells de 2 categorías ha sido evaluado en tres estudios (26, 29, 30) llevados a cabo sobre poblaciones con similar frecuencia de EP (18-20%), con resultados dispares. Los dos estudios con buena calidad (26, 30) muestran resultados compatibles, presentando la escala como más específica que sensible. La sensibilidad en estos estudios oscila entre 60% y 62%.

La rentabilidad diagnóstica de la escala de Wells de 3 categorías ha sido evaluada en 9 de los 11 estudios incluidos en esta revisión. Los estudios realizados sobre población ambulatoria con baja frecuencia de EP (24, 28), a pesar de la similitud en las poblaciones y tener ambos una calidad aceptable, son dispares en cuanto a sus resultados. Así, en el trabajo de Wells (autor que creó y validó la escala anteriormente), la escala presenta una alta sensibilidad (92%) y moderada especificidad; sin embargo, el trabajo de Kline, con criterios de exclusión menos rigurosos aunque con tamaño muestral mayor, presenta una escala con moderadas sensibilidad (54%) y especificidad, siendo mejor esta última. Estos dos trabajos sin embargo aportan resultados similares de valor predictivo negativo de la escala de Wells (99 y 97%, respectivamente).

Con frecuencias moderadas de EP (en torno al 20%) y altas (cercasas o superiores al 30%) los estudios coinciden en valores de la escala de Wells mejores en

sensibilidad que en especificidad. Esto es así pese a las diferencias entre los estudios en cuanto a su calidad (puntuaciones globales que oscilan entre 6 y 13), la composición de su población, la realización de la intervención (diverso grado de experiencia de los médicos que aplican la escala) ó el patrón de referencia usado para la confirmación o descarte del diagnóstico de EP. Sin embargo, estos estudios tienen divergencias en cuanto a la estimación del valor de la sensibilidad de la escala de Wells, que es muy diferente de uno a otro.

En los dos trabajos en los que la población de estudio presenta una moderada frecuencia de EP (29, 30) la escala de Wells muestra una elevada sensibilidad (superior al 90%) y variable especificidad (19% (29) y 63,6% (30)), sin embargo hay que tener en cuenta que uno de estos estudios es el de peor calidad de esta revisión (29).

Cuando la frecuencia de EP es mayor (situada entre 26 y 43%) (23, 25, 31-33) la escala de Wells también se caracteriza por presentar mayor sensibilidad que especificidad, aunque las diferencias entre éstos dos parámetros no son tan marcadas. Los valores de sensibilidad difieren mucho entre los estudios, oscilando en un rango comprendido entre 66% y 91%.

2.2. Escala de Ginebra

En orden de frecuencia, la siguiente escala más estudiada es la de Ginebra, que ha sido valorada en cinco estudios (25, 29, 31-33). Salvo en el estudio de Carrier (29) (el de peor calidad), en todos los trabajos la escala se caracteriza por ser más sensible que específica.

Cabe destacar que sólo uno de los estudios usó la escala de Ginebra de forma prospectiva (25). Si a esto se añade que la calidad de este trabajo era buena, puede tomarse como mejor evidencia disponible la derivada del mismo. En él, con una frecuencia de EP del 26%, la sensibilidad de la escala de Ginebra con empirismo (su forma habitual de uso) fue del 89%, con sólo un 5% de falsos negativos. Esta sensibilidad comparó muy favorablemente con la misma escala usada sin empirismo (72%) y con la escala de Wells (73%).

Los otros tres estudios que valoran la escala de Ginebra fueron realizados en poblaciones con frecuencias de EP entre 29 y 43% (31-33). En todos ellos se calculó de forma retrospectiva, con las limitaciones que esto conlleva. El estudio de Miniati (32), de muy buena calidad, encontró que la escala de Ginebra tenía una sensibilidad muy alta (86%), pero menor que la de Wells y Pisa. En los otros dos estudios, de menor calidad (QUADAS entre 8 y 9 respuestas afirmativas), la sensibilidad de la escala de Ginebra estuvo entre 70% (31) y 78% (33); resultando más baja que la sensibilidad de la escala de Wells en el primer caso y más alta en el segundo.

2.3. Otras escalas

El rendimiento diagnóstico del algoritmo de Wells, la regla de Charlotte y los modelos de Rodger y Pisa ha sido evaluado en un solo estudio para cada uno de ellos.

En un estudio de buena calidad (23), la sensibilidad del algoritmo de Wells se sitúa por encima del 80% y su especificidad está cercana al 30%, resultando una rentabilidad algo superior a la de la escala de Wells, si bien las diferencias no son estadísticamente significativas.

La regla de Charlotte, analizada en el estudio de Kline (28), llevado a cabo sobre población ambulatoria, muestra una alta especificidad (mayor que la de la escala de Wells) y baja sensibilidad (menor que la de la escala de Wells).

El rendimiento diagnóstico del modelo de Rodger resultó algo mayor al de la escala de Wells con dos categorías y a la de Ginebra en el estudio de Carrier (29); sin embargo, la fiabilidad de estos resultados es mínima, debido a los problemas metodológicos de este estudio.

El modelo de Pisa, analizado en un trabajo de muy buena calidad (32), muestra mejor rentabilidad diagnóstica que las escalas de Wells y Ginebra, con diferencias estadísticamente significativas. Así, el área bajo la curva ROC fue del 94% para el modelo de Pisa (frente a 54% en la escala de Ginebra y 75% con la escala de Wells) y el valor predictivo negativo fue del 95% con este modelo (frente a 50% y 88% con las escalas de Ginebra y Wells respectivamente). La única objeción respecto al estudio es en términos de validez externa, ya que se realizó en el mismo centro sanitario en el que la escala de Pisa fue desarrollada y validada.

Discusión

Las evidencias disponibles analizadas en este trabajo, incluyendo estudios de buena calidad, no muestran diferencias relevantes en la rentabilidad diagnóstica de las escalas frente al empirismo.

Al valorar el rendimiento de las diversas escalas, se encuentra que la mayoría de ellas han sido escasamente estudiadas (uno o pocos estudios), siendo la escala de Wells de 3 categorías y la escala de Ginebra las que se han evaluado en un mayor número de investigaciones. La primera ha mostrado buena sensibilidad en poblaciones con una frecuencia moderada de embolismo pulmonar (EP), y sensibilidades en un rango amplio (entre 66 y 91%) en estudios con frecuencia alta. La sensibilidad de la escala de Ginebra, según los estudios de mejor calidad, realizados en poblaciones con alta frecuencia de EP, es superior al 86%. Con todo, es importante destacar que en principio no cabría esperar que la sensibilidad y especificidad de las escalas se vieran afectadas por la prevalencia de embolismo en la población donde se usan (a diferencia de los valores predictivos, que sí son muy sensibles a la frecuencia de enfermedad).

La información aportada por esta revisión sistemática puede ser útil para ayudar a la decisión clínica de usar o no un procedimiento estructurado para calcular la probabilidad pre-test en el diagnóstico del EP. Asimismo, en aquellos casos donde se prefiera utilizar este tipo de procedimientos, puede aportar datos útiles para la elección de una escala concreta según el ámbito del diagnóstico.

> Limitaciones y aspectos metodológicos

Inicialmente se planteó una revisión bibliográfica que fuera lo suficientemente sensible como para recoger la mayor parte de los trabajos en los que se utilizara cualquier método estructurado para la valoración de la probabilidad pretest en el diagnóstico del EP, para lo que se diseñaron amplias estrategias de búsqueda. Si bien esto garantizaría la inclusión de los estudios publicados que abordan algún aspecto relacionado con la pregunta de investigación, como en cualquier revisión sistemática las conclusiones vienen limitadas por la realidad de los trabajos hallados. En este caso concreto, se trata de un conjunto de trabajos muy heterogéneo para el que resulta muy difícil sintetizar la información. En respuesta a esta limitación, se ha optado por una exhaustiva descripción de los estudios, incluyendo el contexto en el que han sido realizados, los criterios en los que se ha basado el análisis de la calidad de los mismos y sus resultados. De esta forma, el potencial usuario de la revisión podrá contrastar la información con sus propios datos o experiencias.

Otro aspecto que conviene no olvidar en cuanto a la realidad de los estudios encontrados, es que todos ellos se cifren a la información sobre el rendimiento diagnóstico de las escalas, pero no aportan datos sobre los resultados que son más relevantes desde el punto de vista clínico. Es decir, aún cuando se encuentra que una escala tiene una alta sensibilidad, no puede derivarse de ello un impacto en términos de menor mortalidad o

más rápida recuperación.

Siguiendo en esta misma línea de una revisión amplia y “sensible”, la calidad de los estudios no ha constituido un criterio de inclusión ó exclusión para la selección. El resultado ha sido una selección en la que se mezclan trabajos de muy buena calidad, como los de Sanson (23) y Miniati (32), con otros como los de Carrier (29) y Ollenberger (33), que aún siendo de mala calidad, aportan algún aspecto adicional en el análisis de la rentabilidad diagnóstica de las escalas. En cualquier caso, los resultados de la revisión de los diferentes trabajos se matizan en función de la calidad de los mismos, valorada mediante la herramienta QUADAS (22), que ha sido validada en este sentido y permite analizar aspectos relacionados con la validez externa e interna de los trabajos.

Sin embargo, la cumplimentación del QUADAS se hace en base a los datos encontrados en las publicaciones, por lo que no puede descartarse que en algunos casos se esté infravalorando la calidad de un estudio por el hecho de no hacer explícitas algunas de sus características.

Para facilitar la síntesis de los resultados los estudios, éstos se han agrupado en función de la forma de determinar la probabilidad pretest y de la frecuencia de embolismo pulmonar en la población de estudio. Sin embargo, para hacer una correcta interpretación de los resultados, es necesario considerar que la frecuencia de diagnóstico depende, además de las características de la población de estudio, de las pruebas de referencia a través de las que se establece el diagnóstico confirmado de EP (23). De este modo, comparar estudios con incidencias similares de EP, no siempre implica que se comparen resultados en poblaciones relativamente semejantes.

Otra consecuencia de la heterogeneidad en las poblaciones y en las intervenciones realizadas en cada uno de los estudios revisados es la falta de un sustrato adecuado para la realización de una síntesis cuantitativa de los resultados.

> **Discusión de los resultados principales**

Los resultados de esta revisión sistemática coinciden con las conclusiones obtenidas en revisiones previas (11, 20,21), en el sentido de no hallar diferencias significativas entre el rendimiento diagnóstico de las escalas frente al empirismo, ni una clara superioridad de una determinada escala frente al resto. Esto es así, a pesar de que esta revisión actualiza a las previas, añadiendo los resultados de diversos estudios realizados en ámbitos diferentes a aquellos en los que las escalas se desarrollaron y validaron inicialmente.

Con respecto a la rentabilidad diagnóstica de las escalas comparada con la del empirismo, los trabajos de buena calidad han mostrado que no existen diferencias o éstas no son significativas desde un punto de vista estadístico. Sólo hay un estudio en el que se ha encontrado diferencia estadísticamente significativa entre el empirismo y una de las escalas en uno de los parámetros de rendimiento diagnóstico valorados; por tanto, se trata de una evidencia muy limitada y su valor desde el punto de vista clínico es pequeño.

Según los autores del grupo Christopher Study Investigators, el principal reto en el desarrollo del diagnóstico del paciente con sospecha de embolismo pulmonar es distinguir de forma rápida y exacta a los pacientes que presentan realmente la enfermedad (que ellos estiman en un 25%) (26). Para lograrlo, la estrategia de cribado ideal debería ser rápida, barata, exacta, fácil de utilizar y con capacidad para excluir el EP (28). En este sentido, asumiendo que el rendimiento diagnóstico de escalas y empirismo fuese similar, el método de elección podría ser el empirismo.

Sin embargo, a la utilización del empirismo como método para determinar la probabilidad pre-test se le atribuye algunos problemas que pueden ser importantes. En primer lugar, la estimación empírica de esta probabilidad tiende a asignar a la mayoría de los pacientes en la categoría de probabilidad intermedia (23), que es la menos útil en la toma de decisiones en el proceso diagnóstico (34). Por otro lado, en este proceso se ha puesto de manifiesto una baja concordancia inter-observadores (12), y el nivel de experiencia del clínico puede influir en la rentabilidad diagnóstica del empirismo (13, 34). Por todo ello, parece conveniente valorar como una alternativa el uso de procedimientos estructurados.

A las escalas o métodos estructurados se les presupone una menor variabilidad inter-observador (23), lo que les confiere mayor consistencia y reproducibilidad, como muestra un estudio de Rodger y Wells (35). En este estudio comparan la fiabilidad inter-observador en la determinación de la probabilidad pre-test a través de la escala de Wells (dicotomizada) y el empirismo, encontrando mejores resultados para la mencionada escala. La mayor fiabilidad de las escalas, junto a la facilidad en su aprendizaje y manejo hace que varios autores recomienden su uso frente al empirismo, especialmente para aquellos clínicos que se encuentran en período de formación o los que tienen poca experiencia en el manejo diagnóstico del EP (11, 12).

En cuanto al modelo o escala a escoger, con la evidencia disponible no puede hablarse de inequívoca superioridad en la rentabilidad diagnóstica de una con respecto a las demás. Por ello, en la elección del modelo influirán decisivamente otros factores, como el tipo de población o la disponibilidad de los datos clínicos y analíticos requeridos para el cálculo de los diferentes modelos, como ya apuntaban Chunilal et al en las conclusiones de su revisión (11). En este sentido, y a título de ejemplo, Ghali et al comentan en su revisión (13) la limitación del uso del algoritmo de Wells, pese a su buena rentabilidad diagnóstica, debido a la excesiva complejidad de su aplicación.

La escala de Wells es, con diferencia, el modelo más evaluado; posiblemente debido a su facilidad de aplicación y versatilidad. Una ventaja es que se ha valorado en poblaciones muy diferentes (incluyendo pacientes exclusivamente ambulatorios) y con distinta prevalencia de EP a la que se presentaba en la población del estudio original (24, 25, 28, 30). Cabe destacar que es una herramienta sencilla de utilizar y que no requiere resultados de pruebas complementarias para su cálculo, lo que supone una enorme ventaja para su aplicación fuera del ámbito hospitalario.

La variante de 3 categorías de la escalas de Wells se caracteriza por presentar una alta sensibilidad y un alto valor predictivo negativo para la categoría de baja probabilidad pre-test, lo que le confiere la propiedad de excluir el diagnóstico de EP con

relativa seguridad en los pacientes asignados a esta categoría. Con la variante de dos categorías se logra un incremento de la especificidad (26, 30, 35), aumentando así el número de pacientes en los que el embolismo pulmonar podría ser excluido, pero a costa de una pérdida importante en la sensibilidad, que baja hasta el 60%.

La siguiente escala, en términos de número de estudios que la evalúan, es la de Ginebra. Entre sus ventajas está la facilidad de uso, que todas sus variables son objetivables, y que su utilización ideal incluye también la valoración empírica (en los casos en los que existe discrepancia entre la probabilidad pre-test calculada a través de la escala y mediante el juicio clínico, prevalece este último). Su principal inconveniente estriba en la necesidad de la realización de pruebas complementarias, reconociéndose como uno de los factores más limitantes a la hora de su aplicación la disponibilidad de la gasometría (13, 34).

La escala de Ginebra fue desarrollada e inicialmente validada en el contexto de pacientes atendidos en un servicio de emergencias, mostrando una alta sensibilidad (17). Los estudios realizados posteriormente muestran resultados diversos. En algunos casos se ha valorado el rendimiento diagnóstico de esta escala en ámbitos diferentes al de origen, mostrando resultados peores que la validación inicial (26, 31). Esto ha llevado a algunos autores a recomendar la limitación del uso de la escala al contexto de los servicios de urgencias (36). Sin embargo, en un estudio de buena calidad, en el que se mezclaban pacientes ambulatorios y hospitalizados, la escala de Ginebra ha tenido una alta sensibilidad (86%) (32), similar a la encontrada previamente en los servicios de urgencias.

El modelo de Pisa presentó una rentabilidad significativamente superior a las escalas de Wells y Ginebra en un trabajo realizado por los mismos autores y en el mismo contexto en el que ésta se desarrolló (32). Además de éste, se presentan otros problemas de validez externa: la inclusión de más variables en el modelo predictivo (sobremodelaje) origina que el modelo se convierta en un reflejo de la población original y consecuentemente sea menos aplicable a una nueva población, lo que conduce a resultados menos favorables cuando el modelo es validado (36). A esta limitación, se añade la dificultad de ser engorroso en el cálculo (37), ya que estima la probabilidad a través del sumatorio de los coeficientes de correlación, cuyo resultado debe ser transportado a una curva de probabilidad.

Con respecto a los demás modelos analizados, el modelo de Rodger y la regla de Charlotte, la escasez de evidencias encontradas y los problemas de validez interna (29) y externa (28) de los estudios que las sostienen no permiten, a priori, recomendar su utilización en ámbitos diferentes a aquellos en los que fueron desarrolladas y validadas inicialmente.

> Implicaciones para la práctica clínica y la investigación

La determinación de la probabilidad pretest para el diagnóstico del embolismo pulmonar, ya sea estimada a través del empirismo ó a través de escalas ó modelos estructurados presenta dos utilidades. Por un lado, clasifica a los pacientes con sospecha de embolismo

pulmonar en categorías de probabilidad que permiten, de forma previa a la realización de pruebas invasivas, descartar el diagnóstico cuando la probabilidad es baja. Por otro lado, cuando ya se ha iniciado el proceso diagnóstico mediante pruebas invasivas, la probabilidad determinada puede orientar en la interpretación de los resultados de dichas pruebas. De hecho, en algunos de los estudios analizados (24, 25, 29, 30), la probabilidad pretest forma parte de los diversos algoritmos de diagnóstico como un elemento de decisión.

En cuanto a la elección del método para la determinación de la probabilidad pretest, suponiendo que la rentabilidad diagnóstica de empirismo y escalas sea equiparable, la utilización de éstas presenta la ventaja de estandarizar y hacer explícito el proceso de valoración de la probabilidad clínica del embolismo pulmonar, así como su resultado.

Con respecto a la decisión de qué escala utilizar, en primer lugar habría que considerar descartar aquellas sobre las que existe poca evidencia. Las evidencias encontradas sobre el rendimiento del algoritmo de Wells, la regla de Charlotte, el modelo de Rodger y el modelo de Pisa se limitan a un estudio para cada uno de ellos, por lo que no puede recomendarse su uso extendido. A la limitación de la escasez de estudios, se suman los problemas, ya comentados, de validez de algunos de estos trabajos, lo que justifica aún más la restricción en las recomendaciones de uso.

Lo recomendado es escoger entre aquellas escalas sobre las que se dispone de más evidencias, más aún, cuando estas evidencias son positivas a favor de su rendimiento. Es el caso de las escalas de Wells de 3 categorías y de Ginebra. En la mayoría de los trabajos de buena calidad que las han evaluado, ambas han mostrado un alto rendimiento diagnóstico, sobre todo a expensas de una alta sensibilidad. Su alto valor predictivo negativo para la categoría de baja probabilidad permite descartar el diagnóstico de embolismo pulmonar con aceptable seguridad.

Además de la rentabilidad diagnóstica, para considerar la elección de una escala sobre otra habrá que tener en cuenta aspectos como la facilidad de su aplicación, el ámbito en el que se aplica (pacientes ambulatorios y/o hospitalizados), posibilidad de uso por parte de cualquier clínico, la incidencia de embolismo pulmonar en la población en la que se pretende aplicar o la disponibilidad de pruebas complementarias rutinarias (electrocardiograma, radiografía de tórax). En este sentido, la falta de disponibilidad de pruebas complementarias, fundamentalmente de la gasometría, constituye una evidente limitación para el uso de la escala de Ginebra. En ámbitos donde esto sucede, como en las consultas de atención primaria, la escala de Wells sería una alternativa válida.

Con independencia del método elegido para la determinación de la probabilidad pretest en el diagnóstico del EP, la exclusión del diagnóstico de embolismo pulmonar no suele basarse exclusivamente en el resultado de esta determinación, sino más bien en el resultado de una combinación de pruebas, de forma que la seguridad se vea reforzada. En este sentido, en muchos estudios, entre ellos algunos de los seleccionados para esta revisión (24, 26, 28, 29, 34), se analiza la rentabilidad diagnóstica de las escalas utilizadas conjuntamente con el resultado de alguna prueba poco invasiva, habitualmente el dímero D u otros parámetros biológicos.

Entre las implicaciones para la investigación se recogen varios aspectos. Con respecto al

análisis del rendimiento de las escalas frente al empirismo, parece que se dispone de suficiente evidencia, por lo que no parece que sean necesarios más estudios en este sentido. Con respecto a la validación de las diversas escalas en ámbitos concretos en los que se pretende que sean aplicadas, sería conveniente considerar la importancia de ajustarse al formato original de las escalas. De otro modo, será complicada la comparación con los resultados de otros estudios y, de esta forma, su validación. En este sentido es necesario recalcar la importancia de hacer estudios prospectivos, fundamentalmente con aquellas escalas que incorporan algún elemento subjetivo entre sus componentes, como la escala de Wells ó en su forma de ser aplicada, como la escala de Ginebra. Por último, considerar la necesidad de estudiar otros resultados que vayan más allá del rendimiento diagnóstico. Son necesarios estudios sobre la efectividad clínica de las escalas (con resultados del tipo mortalidad, complicaciones, efecto sobre la duración de la estancia hospitalaria) analizadas de forma independiente a otros componentes de los algoritmos diagnósticos para el embolismo pulmonar.

Conclusiones

1. La rentabilidad diagnóstica de las escalas utilizadas para la determinación de la probabilidad pretest en el diagnóstico del embolismo pulmonar no presenta diferencias relevantes con la rentabilidad proporcionada por la valoración empírica. Esta evidencia se ha constatado en al menos tres estudios de buena calidad.
2. El análisis de la evidencia disponible no ha puesto de manifiesto de forma clara la superioridad de la rentabilidad diagnóstica de una escala concreta sobre las demás.
3. Las escalas de Wells de tres categorías y Ginebra son las que se han evaluado en un mayor número de estudios, además de haberse valorado su rendimiento diagnóstico en ámbitos diferentes a aquellos en los que fueron desarrolladas y validadas.
4. La escala de Wells de tres categorías ha mostrado una alta sensibilidad (mayor al 90%) en los dos estudios que la han valorado en poblaciones con una frecuencia moderada de embolismo pulmonar (en torno al 20%), si bien la especificidad fue baja (menor del 65%). Los resultados sobre la sensibilidad de esta escala en los estudios realizados en poblaciones con alta frecuencia de embolismo pulmonar (entre 26 y 43%) son discordantes, oscilando entre el 66 y 91%.
5. Los dos estudios de mejor calidad de todos los que valoran la escala de Ginebra, realizados en poblaciones con alta frecuencia de embolismo pulmonar, coinciden en calificar su sensibilidad como alta (mayor de 86%).
6. Los estudios sobre rentabilidad diagnóstica hallados en esta revisión no han valorado la efectividad clínica de las escalas en términos de resultados en salud.

REFERENCIAS

1. Uresandi F, Blanquer J, Conget F, de Gregorio MA, Lobo JL, Otero R, et al. Recommendations of the Spanish Society of Pulmonology and Thoracic Surgery (SEPAR). Guidelines for the diagnosis, treatment, and follow up of pulmonary embolism. *Arch Bronconeumol*. 2004;40:580-594.
2. Task Force on Pulmonary Embolism. European Society of Cardiology. Guidelines on diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2000;21:1301-1336.
3. British Thoracic Society. Suspected acute pulmonary embolism: a practical approach. *Thorax*. 1997;52:S1-24.
4. British Thoracic Society Standards of Care Committee Pulmonary Embolism Guideline Development Group. British Thoracic Society guidelines for the management of suspected acute pulmonary embolism. *Thorax*. 2003;58:470-484.
5. Sáenz C, Sánchez V, Velázquez MT, Tello R, Gómez MA, Delgado J et al. Guías de práctica clínica de la Sociedad Española de Cardiología en tromboembolismo e hipertensión pulmonar. *Rev Esp Cardiol*. 2001;54:194-210.
6. PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. *JAMA*. 1990;263:2753-2759.
7. Worsley D, Alavi A. Comprehensive análisis of the results of the PIOPED study. *J Nucl Med*. 1995;36:2380-2387.
8. Perrier A. Noninvasive diagnosis of pulmonary embolism. *Haematologica*. 1997; 82:328-331.
9. Miniati M, Monti S, Bottai M. A structured clinical model for predicting the probability of pulmonary embolism. *Am J Med*. 2003;114:173-179.
10. Pini M, Marchini L, Giordano A. Diagnostic strategies in venous thromboembolism. *Haematologica*. 1999;84:535-540.
11. Chunilal SD, Eikelboom JW, Attia J, Miniati M, Panju AA, Simel DL, et al. Does this patient have pulmonary embolism?. *JAMA*. 2003;290:2849-2858.
12. Kelly J, Hunt BJ. The utility of pretest probability assessment in patients with clinically suspected venous thromboembolism. *J Thromb Haemost*. 2003;1:1888-1896.
13. Ghali WA, Cornuz J, Perrier A. New methods for estimating pretest probability in the diagnosis of pulmonary embolism. *Curr Opin Pulm Med*. 2001;7:349-353.

14. Runyon m, Webb W, Jones A, Kline J. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the canadian Score and the Charlotte Rule: a prospective observational study. *Acad Emerg Med.* 2005;12:587-593.
15. Wells PS, Ginsberg JS, Anderson DR, Kearon C, Gent M, Turpie AG et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med.* 1998;129:997-1005.
16. Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. *Thromb Haemost.* 2000;83:416-420.
17. Wicki J, Perneger TV, Junod AF, Bounameaux H, Perrier A. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. *Arch Intern Med.* 2001;161:92-97.
18. Kline JA, Nelson RD, Jackson RE, Courtney DM. Criteria for the safe use of D-dimer testing in emergency department patients with suspected pulmonary embolism: A multicenter US study. *Ann Emerg Med.* 2002;39:144-152.
19. Rodger MA, Raymond F, Stiell I, Nair R, Bredeson CN, Jones G, et al. Development and validation of a simple clinical prediction rule to exclude pulmonary embolism. *Blood.* 1999;94:27a.
20. Tamariz LJ, Eng J, Segal JB, Krishnan JA, Bolger DT, Streiff MB et al. Usefulness of clinical prediction rules for the diagnosis of venous thromboembolism: A systematic review. *Am J Med.* 2004;117:676-684.
21. Segal JB, Eng J, Tamariz LJ, Bass EB. Review of the evidence on diagnosis of deep venous thrombosis and pulmonary embolism. *Ann Fam Med.* 2007;5:63-73.
22. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25. Available from: <http://www.biomedcentral.com/content/pdf/1471-2288-3-25.pdf>.
23. Sanson BJ, Lijmer JG, Mac Gillavry MR, Turkstra F, Prins MH, Buller HR. Comparison of a clinical probability estimate and two clinical models in patients with suspected pulmonary embolism. ANTELOPE-Study Group. *Thromb Haemost.* 2000;83:199-203.
24. Wells PS, Anderson DR, Rodger M, Stiell I, Dreyer JF, Barnes D et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med.* 2001;135:98-107.

25. Chagnon I, Bounameaux H, Aujesky D, Roy PM, Gourdier AL, Cornuz J et al. Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism. *Am J Med.* 2002;113:269-275.
26. Huisman MV. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA.* 2006;295:172-179.
27. Miniati M, Monti S, Bauleo C, Scoscia E, Tonelli L, Dainelli A et al. A diagnostic strategy for pulmonary embolism based on standardised pretest probability and perfusion lung scanning: a management study. *Eur J Nucl Med Mol Imaging.* 2003; 30:1450-1456.
28. Kline JA, Runyon MS, Webb WB, Jones AE, Mitchell AM. Prospective study of the diagnostic accuracy of the simplify D-dimer assay for pulmonary embolism in emergency department patients. *Chest.* 2006;129:1417-1423.
29. Carrier M, Wells PS, Rodger MA. Excluding pulmonary embolism at the bedside with low pre-test probability and D-dimer: Safety and clinical utility of 4 methods to assign pre-test probability. *Thromb Res.* 2006;117:469-474.
30. Penaloza A, Melot C, Dochy E, Blocklet D, Gevenois PA, Wautrecht JC et al. Assessment of pretest probability of pulmonary embolism in the emergency department by physicians in training using the Wells model. *Thromb Res.* 2007; 120:173-179.
31. Moores LK, Collen JF, Woods KM, Shorr AF. Practical utility of clinical prediction rules for suspected acute pulmonary embolism in a large academic institution. *Thromb Res.* 2004;113:1-6.
32. Miniati M, Bottai M, Monti S. Comparison of 3 clinical models for predicting the probability of pulmonary embolism. *Medicine.* 2005;84:107-114.
33. Ollenberger GP, Worsley DF. Effect of patient location on the performance of clinical models to predict pulmonary embolism. *Thromb Res.* 2006;118:685-690.
34. Kline JA, Wells PS. Methodology for a rapid protocol to rule out pulmonary embolism in the emergency department. *Ann Emerg Med.* 2003;42:266-275.
35. Rodger MA, Maser E, Stiell I, Howley HE, Wells PS. The interobserver reliability of pretest probability assessment in patients with suspected pulmonary embolism. *Thromb Res.* 2005;116:101-107.
36. Righini M, Bounameaux H. External validation and comparison of recently described prediction rules for suspected pulmonary embolism. *Curr Op*

Pulm Med. 2004;10:345-349.

37. Alonso Martinez JL, Garcia Sanchotena JL, Abinzano Guillen ML, Urbieto Echezarreta MA, Anniccherico Sanchez FJ, Fernández LV. Utilidad de los modelos clínicos en la predicción de tromboembolia pulmonar. An Med Interna. 2006;23:264-268.

ANEXO I. Cuestionario QUADAS

Cuestionario QUADAS para la evaluación de la calidad de estudios sobre rendimiento diagnóstico.

ITEM	PREGUNTA	POSIBLES RESPUESTAS			
		Si	No	No claro	No aplicable
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	X	X	X	
2	Were selection criteria clearly described?	X	X	X	
3	Is the reference standard likely to correctly classify the target condition?	X	X	X	
4	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two test?	X	X	X	X
5	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	X	X	X	X
6	Did the patients receive the same reference standard regardless of the index test result?	X	X	X	
7	Was the reference standard independent of the index test?	X	X	X	X
8	Was the execution of the index test described in sufficient detail to permit replication of the test?	X	X	X	X
9	Was the execution of the reference standard described in sufficient detail to permit its replication?	X	X	X	X
10	Were the index test results interpreted without knowledge of the results of the reference standard?	X	X	X	X
11	Were the reference standard results interpreted without knowledge of the results of the index test?	X	X	X	X
12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	X	X	X	X
13	Were uninterpretable/intermediate test results reported?	X	X	X	
14	Were withdrawals from the study explained?	X	X	X	

Fuente: Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25. Available from: <http://www.biomedcentral.com/content/pdf/1471-2288-3-25.pdf>.

ANEXO II. Exclusiones

Motivo fundamental	Referencia bibliográfica	Detalle del motivo de no selección
El tema no está relacionado con el objetivo de la revisión	Vermeer HJ, Ypma P, Van Strijen MJL, Muradin AA, Hudig F, Jansen RW et al. Exclusion of venous thromboembolism: Evaluation of D-Dimer PLUS for the quantitative determination of D-dimer. <i>Thromb Res.</i> 2005;115:381-386.	Estudian pacientes con TVP y EP sin distinguir entre ambas manifestaciones de la ETEV en los resultados
	Siragusa S, Anastasio R, Porta C, Falaschi F, Pirrelli S, Palmieri P et al. Deferment of objective assessment of deep vein thrombosis and pulmonary embolism without increased risk of thrombosis: a practical approach based on the pretest clinical model, D-dimer testing, and the use of low-molecular-weight heparins. <i>Arch Intern Med.</i> 2004;164:2477-2482.	
	Kovacs MJ, MacKinnon KM, Anderson D, O'Rourke K, Keeney M, Kearon C et al. A comparison of three rapid D-dimer methods for the diagnosis of venous thromboembolism. <i>Br J Haematol.</i> 2001;115:140-144.	
	Rodger MA, Maser E, Stiell I, Howley HE, Wells PS. The interobserver reliability of pretest probability assessment in patients with suspected pulmonary embolism. <i>Thromb Res.</i> 2005;116:101-107.	Estudia la fiabilidad interobservador en la determinación de la PPT sin comparar con el diagnóstico final
	Wolf SJ, McCubbin TR, Feldhaus KM, Faragher JP, Adcock DM. Prospective validation of Wells Criteria in the evaluation of patients with suspected pulmonary embolism. <i>Ann Emerg Med.</i> 2004;44:503-510.	Evalúa aspectos metodológicos en el desarrollo de la escala
	Parent F, Maitre S, Meyer G, Raheison C, Mal H, Lancar R et al. Diagnostic value of D-dimer in patients with suspected pulmonary embolism: Results from a multicentre outcome study. <i>Thromb Res.</i> 2007;120:195-200.	Determinan la probabilidad pretest sin hacer uso de escalas
	Ghanima W, Abdelnoor M, Mowinckel MC, Sandset PM. The performance of STA-Liatest D-dimer assay in out-patients with suspected pulmonary embolism. <i>Br J Haematol.</i> 2006;132:210-215.	
	Nilsson T, Mare K, Carlsson A. Value of structured clinical and scintigraphic protocols in acute pulmonary embolism. <i>J Intern Med.</i> 2001;250:213-218.	
	Rodger MA, Carrier M, Jones GN, Rasuli P, Raymond F, Djunaedi H et al. Diagnostic value of arterial blood gas measurement in suspected pulmonary embolism. <i>Am J Respir Crit Care Med.</i> 2000;162:2105-2108.	
	Perrier A, Desmarais S, Miron MJ, Moerloose P, Lepage R, Slosman D, et al. Non-invasive diagnosis of venous thromboembolism in outpatients. <i>Lancet.</i> 1999; 353:190-95.	Revisión sistemática sobre el diagnóstico del EP que no hace referencia a escalas
Roy PM, Colombet I, Durieux P, Chatellier G, Sors H, Meyer G. Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. <i>BMJ.</i> 2005;331:259-263.		

Motivo fundamental	Referencia bibliográfica	Detalle del motivo de no selección
No es un estudio original ni una revisión sistemática	Hogg KE, Brown MD, Kline JA. Estimating the pretest probability threshold to justify empiric administration of heparin prior to pulmonary vascular imaging for pulmonary embolism. <i>Thromb Res.</i> 2006;118:547-553.	Revisión no sistemática de la literatura sobre el uso de escalas y otros procedimientos diagnósticos en el EP
	Wells PS. Advances in the diagnosis of venous thromboembolism. <i>J Thromb Trombolysis.</i> 2006; 21(1):31-40.	
	Gudino L. Initial clinical suspicion as a key element when deciding the diagnosis of pulmonary embolism. <i>Salud Sci.</i> 2006;15:452-455.	
	Palla A, Bardi G, Ribas C. Diagnosis of pulmonary embolism. <i>Semin Thromb Hemost.</i> 2006;32:822-830.	
	Morris TA, Follett H. New modalities for the diagnosis of deep venous thrombosis and pulmonary embolism. <i>Clin Pulm Med.</i> 2005;12:359-362.	
	Michiels JJ, G�adisseur A, Van Der Planken M, Schroyens W, De Maeseneer M, Hermesen JT et al. A critical appraisal of non-invasive diagnosis and exclusion of deep vein thrombosis and pulmonary embolism in outpatients with suspected deep vein thrombosis or pulmonary embolism: How many tests do we need?. <i>Int Angiol.</i> 2005;24:27-39.	
	Michiels JJ, G�adisseur A, van der PM, Schroyens W, De Maeseneer M, Hermesen JT et al. Screening for deep vein thrombosis and pulmonary embolism in outpatients with suspected DVT or PE by the sequential use of clinical score: a sensitive quantitative D-dimer test and noninvasive diagnostic tools. <i>Seminars in Vascular Medicine</i> 2005;5:351-364.	
	Chu K, Brown AFT. Likelihood ratios increase diagnostic certainty in pulmonary embolism. <i>Emerg Med Australas.</i> 2005; 17:322-329.	
	Righini M, Bounameaux H. External validation and comparison of recently described prediction rules for suspected pulmonary embolism. <i>Curr Opin Pulm Med.</i> 2004;10:345-349.	
	Riedel M. Diagnosing pulmonary embolism. <i>Postgrad Med J.</i> 2004;80:309-319.	
	Brown G, Hogg K. Towards evidence based emergency medicine: best BETs from the Manchester Royal Infirmary. Clinical probability scoring and pulmonary embolism. <i>Emerg Med J.</i> 2003;20:367.	
	Boyd R, Hogg K. Towards evidence based emergency medicine: best BETs from the Manchester Royal Infirmary. Accuracy of combining clinical probability score and simpliRED D-dimer for diagnosis of pulmonary embolism. <i>Emerg Med J.</i> 2003;20:367-368.	
	Michiels JJ, Schroyens W, De Backer W, van der PM, Hoogsteden H, Pattynama PM. Non-invasive exclusion and diagnosis of pulmonary embolism by sequential use of the rapid ELISA D-dimer assay, clinical score and spiral CT. <i>Int Angiol.</i> 2003;22:1-14.	

Motivo fundamental	Referencia bibliográfica	Detalle del motivo de no selección
Diseño no adecuado	Corral GL, Guijo HT, Moreiro BM, Cordero M, Perez Arellano JL, del Pino MJ. Pulmonary embolism: analysis of cases without initial suspicion and sensitivity of three clinical models. Med Clin (Barc) 2003;121:601-605.	Estudio de caso-control
Cumplen otros criterios de exclusión (población, intervención, comparación, resultados)	Runyon MS, Webb WB, Jones AE, Kline JA. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the Canadian score and the Charlotte rule: a prospective observational study. Acad Emerg Med.2005;12:587-593.	Compara empirismo frente a escalas de Wells y Charlotte; no se excluye a la población pediátrica y el seguimiento es de 45 días
	Bosson JL, Barro C, Satger B, Carpentier PH, Polack B, Perron G. Quantitative high D-dimer value is predictive of pulmonary embolism occurrence independently of clinical score in a well-defined low risk factor population. J Thromb Haemost.2005;3:93-99.	Analiza el valor diagnóstico del DD frente a escalas, pero son en pacientes con baja PPT
	Sohne M, Kruij MJ, Nijkeuter M, Tick L, Kwakkel H, Halkes SJ et al. Accuracy of clinical decision rule, D-dimer and spiral computed tomography in patients with malignancy, previous venous thromboembolism, COPD or heart failure and in older patients with suspected pulmonary embolism. J Thromb Haemost.2006;4:1042-1046.	En el estudio sólo se incluyen pacientes con determinadas patologías
	Alonso JL, García JL, Abinzano ML, Urbieto MA, Annicchero FJ, Fernandez L. Utility of the clinical models for predicting pulmonary thromboembolism. An Med Interna.2006;23:264-268.	Estudio sobre pacientes diagnosticados de EP
	Courtney DM, Kline JA. Prospective use of a clinical decision rule to identify pulmonary embolism as likely cause of outpatient cardiac arrest. Resuscitation.2005;65:57-64.	Estudio sobre pacientes que ingresan por síncope
	Perrier A, Roy PM, Sanchez O, LE Gal G, Meyer G, Gourdier AL et al. Multidetector-row computed tomography in suspected pulmonary embolism. New Engl J Med.2005;352:1760-1768.	La determinación de la PPT se realiza a través de escalas y de valoración clínica no estructurada, sin que en los resultados se distinga entre un método y otro
	Righini M, Aujesky D, Roy PM, Cornuz J, De Moerloose P, Bounameaux H et al. Clinical usefulness of D-dimer depending on clinical probability and cutoff value in outpatients with suspected pulmonary embolism 134. Arch Intern Med 2004; 164(22):2483-2487.	
	Perrier A, Roy PM, Aujesky D, Chagnon I, Howarth N, Gourdier A, et al. Diagnosing pulmonary embolism in outpatients with clinical assessment, D-dimer measurement, venous ultrasound, and helical computed tomography: a multicenter management study. Am J Med.2004;116:291-299.	
Hogg K, Dawson D, Mackway-Jones K. The emergency department utility of simplify D-dimer(trademark) to exclude pulmonary embolism in patients with pleuritic chest pain. Ann Emerg Med.2005;46:305-310.	Introducen una modificación en la escala de Wells	

Motivo fundamental	Referencia bibliográfica	Detalle del motivo de no selección
No es un estudio original ni una revisión sistemática (continuación)	Mountain D. Diagnosing pulmonary embolism: A question of too much choice? Emerg Med.2003; 15(3):250-262.	Revisión no sistemática de la literatura sobre el uso de escalas y otros procedimientos diagnósticos en el EP
	Mcginn T, Moore C, Ho W. Practice corner: Using clinical prediction rules. Evid Based Med.2002;7:132-133.	
	Ghali WA, Comuz J, Perrier A. New methods for estimating pretest probability in the diagnosis of pulmonary embolism. Curr Opin Pulm Med.2001;7:349-353.	
	Wolfe TR, Hartsell SC. Pulmonary embolism: making sense of the diagnostic evaluation. Ann Emerg Med.2001;37:504-514.	
	Michiels JJ, Pattynama PM. Exclusion and diagnosis of pulmonary embolism by a rapid ELISA D-dimer test and noninvasive imaging techniques within the context of a clinical model. Clin Appl Thromb Hemost.2000;6:46-52.	
	Pini M, Marchini L, Giordano A. Diagnostic strategies in venous thromboembolism. Haematologica.1999;84:535-540.	
	Perrier A. Noninvasive diagnosis of pulmonary embolism. Haematologica.1997;82:328-331.	
	Kelly J, Hunt BJ. The utility of pretest probability assessment in patients with clinically suspected venous thromboembolism. J Thromb Haemost.2003;1:1888-1896.	Carta
	Leclercq MGL, Kruij MJHA, Mac GillavryMR, Marwijk Kooy M, Buller HR. Observer variability in the assessment of clinical probability in patients with suspected pulmonary embolism. J Thromb Haemost.2004;2:1204-1206.	Protocolo de manejo en urgencias
	Kline JA, Wells PS. Methodology for a rapid protocol to rule out pulmonary embolism in the emergency department. Ann Emerg Med.2003;42:266-275.	Revisiones de artículos originales
	Damek HA, Jr., Lindbloom EJ. How accurate is the D-dimer assay in diagnosing pulmonary embolism? J Fam Pract.2002;51:919.	
	Validated algorithm for suspected pulmonary embolism. J Fam Pract.2005;54:653-657.	
	D-dimer useful for excluding deep vein thrombosis and pulmonary embolism. J Fam Pract.2004;53:686-688.	

Motivo fundamental	Referencia bibliográfica	Detalle del motivo de no selección
Son trabajos sobre el desarrollo y/o validación de escalas ó modelos	Michel BC, Kuijter PM, McDonnell J, van Beek EJ, Rutten FF, Buller HR. The role of a decision rule in symptomatic pulmonary embolism patients with a non-high probability ventilation-perfusion scan. <i>Thromb Haemost.</i> 1997;78:794-798.	Desarrollo de un modelo sin validación
	Wells PS, Ginsberg JS, Anderson DR, Kearon C, Gent M, Turpie AG et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. <i>Ann Intern Med.</i> 1998;129:997-1005.	Desarrollo y validación de un modelo: Algoritmo de Wells
	Miniati M, Prediletto R, Formichi B, Marini C, Di Ricco G, Tonelli L, et al. Accuracy of clinical assessment in the diagnosis of pulmonary embolism. <i>Am J Respir Crit Care Med.</i> 1999;159:864-871.	Desarrollo y validación de un modelo
	Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. <i>Thromb Haemost.</i> 2000;83:416-420.	Desarrollo y validación de un modelo: Escala de Wells
	Stollberger C, Finsterer J, Lutz W, Stoiberl C, Kroiss A, Valentin A et al. Multivariate analysis-based prediction rule for pulmonary embolism. <i>Thromb Res.</i> 2000;97:267-273.	Desarrollo y validación de un modelo (Austria)
	Leveau P, Ingrand P. Clinical probability of the diagnosis of pulmonary embolism calculated by the ASE score. <i>Réanim Urgences.</i> 2000;9:101-108.	Desarrollo y validación de un modelo: Escala ASE (Francia)
	Wicki J, Perneger TV, Junod AF, Bounameaux H, Perrier A. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. <i>Arch Intern Med.</i> 2001;161:92-97.	Desarrollo y validación de un modelo: Escala de Ginebra
	Kline JA, Nelson RD, Jackson RE, Courtney DM. Criteria for the safe use of D-dimer testing in emergency department patients with suspected pulmonary embolism: A multicenter US study. <i>Ann Emerg Med.</i> 2002;39:144-152.	Desarrollo de un modelo: Regla de Charlotte junto al dímero D
	Miniati M, Monti S, Bottai M. A structured clinical model for predicting the probability of pulmonary embolism. <i>Am J Med.</i> 2003;114:173-179.	Desarrollo y validación de un modelo: Modelo de Pisa
	Kline JA, Mitchell AM, Kabrhel C, Richman PB, Courtney DM. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. <i>J Thromb Haemost.</i> 2004;2:1247-1255.	Desarrollo y validación de un modelo: Regla de Charlotte
Palud L, Laurent M, Gueret P, Meunier C, Garin E, Benoit PO et al. Value of the association of D-dimer measurement and the evaluation of clinical probability in a non-invasive diagnostic strategy of pulmonary embolism. <i>Arch Mal Coeur Vaiss.</i> 2004;97:93-99.	Desarrollo y validación de un modelo (Francia)	
Chen JY, Chao TH, Guo YL, Hsu CH, Huang YY, Chen JH et al. A simplified clinical model to predict pulmonary embolism in patients with acute dyspnea. <i>Int Heart J.</i> 2006;47:259-271.	Desarrollo y validación de un modelo (Taiwan)	
Le Gal G, Righini M, Roy PM, Sanchez O, Aujesky D, Bounameaux H et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. <i>Ann Intern Med.</i> 2006;144:165-171.	Desarrollo y validación de un modelo: Escala de Ginebra revisada	

**AGENCIA DE EVALUACIÓN DE TECNOLOGÍAS
SANITARIAS DE ANDALUCÍA (AETSA)**

Avda de la Innovación s/n. Edificio Renta Sevilla. 2ª planta
41020 Sevilla. España (Spain)
Tlf. 955 00 68 04 / Fax. 955 00 68 45

