SUBMITTED 14 AUG 23
REVISION REQ. 25 OCT 23; REVISION RECD. 30 NOV 23
ACCEPTED 19 DEC 23
ONLINE-FIRST: JANUARY 2024
DOI: https://doi.org/10.18295/squmj.1.2024.007

Diagnostic Test Accuracy of the YEARS Algorithm for Pulmonary Embolism

A systematic review and meta-analysis

***Sean R. Te Haara,[1] Helena De Rezende,[2] Chao Wang[3]**

[1]Specialist Assets and Resilience, London Ambulance Service, London, United Kingdom; [2]Department of Nursing Science, Bournemouth University, Bournemouth, United Kingdom; [3]Centre for Health and Social Care Research, School of Education, Midwifery and Social Work, Kingston University, Kingston upon Thames, United Kingdom.
**Corresponding Author's e-mail: stehaara@outlook.com*

Abstract

This systematic review and meta-analysis aimed to evaluate both the diagnostic test accuracy of the YEARS algorithm in excluding pulmonary embolism and to compare the advance imaging utilisation rate of YEARS against standard practice. Published studies were selected across several databases from July 2017 to September 2022 using Joanna Briggs Institute methodology for systematic reviews of diagnostic accuracy. The analysis included ten studies with nearly 14,000 participants. YEARS showed a sensitivity of 96% (95% CI 93-98%) and specificity of 50% (95% CI 33-67%). The risk ratio for advanced imaging was 0.78 (95% CI 67-90), showing an overall reduction. YEARS is an effective means of safely managing patients with suspected pulmonary embolism.

Keywords: Pulmonary Embolism; Diagnostic Imaging; Fibrin Fragment D; Meta-Analysis; Systematic Review; Fibrin Fibrinogen Degradation Products; Clinical Decision Rules.

Introduction

Pulmonary Embolism (PE) carries an incidence between 39-115 per 100,000 people and is the third most common cardiovascular cause of death worldwide.[1,2] The presentation of PE, though sometimes causing overt cardiovascular compromise, can often be non-specific. For this reason, PE is not uncommon yet can be difficult to diagnose leading to over-

investigation, inappropriate diagnosis, prolonged use of Emergency Department (ED) resources and inappropriate treatment.[3-5] Evidence shows a recent rise in the use of emergency resources to investigate PE compared to previous decades.[6,7] Despite this increase in investigation and diagnosis of PE, the overall mortality has not significantly changed.[3,8]

To help with diagnosing patients with suspected PE whilst avoiding the unnecessary use of Advanced Imaging (AI), several algorithms have been developed. The YEARS algorithm was first published in the Lancet on May 2017.[9] The algorithm used an abbreviated version of the WELLS algorithm consisting of three criteria: clinical signs and symptoms of Deep Vein Thrombosis (DVT), likely diagnosis of PE, and the presence of haemoptysis. Patients are classified as either having a low (no criteria present) or high (one or more criteria present) pre-test probability of having PE. Based on this pre-test probability, a lower or higher D-dimer decision threshold, 500 ng/ml or 1000 ng/ml respectively, is used to categorize patients as low risk (not requiring AI) or high risk (requiring AI). YEARS is presented in Figure 1. The algorithm had a sensitivity of 98%, specificity of 55%, positive predictive value 25%, negative predictive value 99%.[9]

The gold standard in AI for diagnosing PE is Computed Tomography Pulmonary Angiography (CTPA).[10,11] This method carries several risks to the patient such as cancer development secondary to radiological exposure,[12] as well as the risk of nephrotoxicity and anaphylaxis from the required intravenous contrast.[13] Additionally, Computer Tomography (CT) scans are not always readily available and a reduction in their use would likely lead to budgetary savings and better use of emergency resources.[14] Alternative AI modalities exist which can modify patient risk but do not negate them entirely.[15] This calls into question the liberal use of AI on patients with a low or insignificant chance of having PE.[11] Thus, we pose the following question, what is the diagnostic test accuracy and utility of the YEARS algorithm at excluding PE?

A preliminary search on CINAHL Plus and Medline revealed one systematic review on this topic. The review analysed four different algorithms (one of which being YEARS) within four specific patient sub-groups.[16] Furthermore, this previously published review retrospectively implemented its own study protocol on cohorts from various studies and compared three different clinical decision rules to YEARS.[16] In light of this evidence, we aimed to evaluate the accuracy of YEARS in excluding PE and compare YEARS' AI utilisation rate against standard practice. Our review proposes to broaden this theme as our

included studies differed in our cohort population and study design. Furthermore, our outcome metrics differed from their singular use of missed PE and included sensitivity, specificity, likelihood ratios and predictive values.

When evaluating an algorithm designed to exclude PE whilst negating AI use, patient safety is fundamental. This will be demonstrated by estimating both the likelihood of missed PE and the incidence of AI exposure. When analysing the utilisation rate of AI, a reference test will be used to demonstrate either an increase or decrease. The reference test will be termed standard practice and may include any alternative algorithms for investigating suspected PE (for example the WELLS algorithm). The objectives of this study are to evaluate the accuracy of YEARS in excluding PE defined as the sensitivity, specificity, likelihood ratios and predictive values. Additionally, to compare the AI utilisation rate of YEARS against standard practice via calculation of the associated risk ratio.

Methods

We followed the Joanna Briggs Institute (JBI) methodology for systematic reviews of diagnostic test accuracy.[17] Moreover, the study was reported using the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) guidelines and the PRISMA of diagnostic test accuracy as an extension.[18,19]

Inclusion Criteria

The Population, Index test, Reference test and Diagnosis of interest (PIRD) model was utilised to develop the inclusion and exclusion criteria.[17,19]

Population: Suspected PE in individuals aged 16 years or older of any ethnicity or gender; living in any geographical location. Studies conducted in ED, inpatient, and outpatient departments were included. Studies involving pregnant participants were excluded.

Index Test: The original YEARS algorithm with D-dimer decision thresholds of 500 ng/ml or 1000 ng/ml respectively depending on YEARS score.

Reference Test: A reference test is required to detect true/false PE. This will include either unilateral validation via AI or alternatively, prospective implementation of YEARS with three-month participant follow up if no AI is ordered (as is commonly utilised in the Venous

Thromboembolism [VTE] literature). The use of alternative PE algorithms were used as a reference test to assess the rate of AI utilisation.

Diagnosis of Interest: Any thromboembolism within the pulmonary circulatory tree as diagnosed on AI/autopsy. This review shall not delineate between Isolated Sub-segmental Pulmonary Embolism (ISPE) and PE to improve homogeneity across all studies. Subsequently, DVT (without PE) will not be counted as missed PE.

Study Design: randomised control trials, case-controlled studies, cross-sectional studies, and retrospective or prospective cohort studies published in peer reviewed journals were included. Studies which conducted post-hoc analysis upon literature already included for review was excluded to avoid repetition of synthesised outcomes.

Search Strategy

The search strategy was developed with the input of the librarian and the following terms were applied to the databases: (“pulmonary embol*” OR “pulmonary thromboembol*” OR "fibrin split product*" OR "Fibrin degradation product*" OR "D-dimer") [TI, AB] AND (“years score” OR “years study” OR “years algorithm” OR “years tool” OR “years criteria” OR “years rule” OR “years clinical decision” OR “years diagnostic”) [TX]

An initial limited search (Step 1) was performed in Medline (EBSCO host, Birmingham, Alabama, USA) in order to identify key words and indexed terms. This informed the development of a comprehensive search strategy (listed above) which was adapted for each database searched (Step 2). The databases used to collate studies were CINAHL Plus (EBSCO host, Birmingham, Alabama, USA), AMED (EBSCO host, Birmingham, Alabama, USA), Medline (EBSCO host, Birmingham, Alabama, USA), and EMBASE (Ovid Technologies, Inc., New York, USA).

All studies published in English between July 2017 (the original publication date of the YEARS algorithm) to September 2022 (date of database search) were included.[9] Data was managed using Rayyan™ reference manager and Microsoft Excel Software™ [20,21] and was assessed by two independent authors. Disagreements were resolved through discussion.

Assessment of Methodological Quality

Selected studies were critically appraised for methodological quality using the JBI checklist for diagnostic test accuracy studies Critical Appraisal Tool – CAT. This instrument is based on 10 ‘‘signalling questions’’ from the revised Quality Assessment of Diagnostic Studies (QUADAS-2) critical appraisal tool for diagnostic test accuracy.[22] The four domains used to assess risk of bias included patient selection (questions 1 to 3), index tests (questions 4 and 5), reference standard (questions 6 and 7), and flow and timing (questions 8 to 10). This provided objective appraisal of potential bias present within the included studies. All studies were included regardless of methodological quality and 20% of the total included studies were quality assured between two authors with the remaining critiqued via a single researcher.

Data Extraction

Data from selected studies was extracted by one author utilising a modified version of the Standards for Reporting Diagnostic accuracy studies (STARD) checklist to fit this reviews aims and objectives.[23] Two authors completed independent data extraction of a minimum 20% studies as quality assurance. Upon data collection, if required data was not made available we calculated the values from what was provided. If this was not possible authors of included studies were contacted to provide additional data.

Data Analysis and Synthesis

Data was synthesized using meta-analysis as described below. When meta-analysis was not performed the synthesis without meta-analysis reporting guidelines were utilised.[24] We have performed meta-analysis of the test accuracy in terms of sensitivity and specificity as per the JBI methodology for systematic reviews of diagnostic test accuracy.[17] The meta-analysis results have been presented in a forest plot and summary receiver operating characterises (ROC) curve.[19,24,25] A hierarchical random-effects logit model was employed using Stata package metadta.[26] In addition, we were interested in the impact of YEARS in terms of reducing scans compared to other algorithms. We collected data from studies and performed a meta-analysis of the risk ratio of YEARS incurring imaging. Profile likelihood method was adopted as suggested by Kontopantelis and Reeves,[27] using the ‘metan’ package in Stata 18.[28] The result was again presented in a forest plot. In addition, subgroup analysis was performed, to assess the heterogeneity between study design (prospective and retrospective) regarding both outcomes: diagnostic test accuracy and impact on AI utilisation. Statistical analysis mirrored the protocols described above and was presented via forest plots. This was

reasoned to be important as the difference in study design may be a significant contributing factor of heterogeneity between articles.

Results

Study Inclusion

In total, 226 studies were retrieved and abstracts read. 25 papers were assessed for full-text reading, for which 15 were excluded with reasons. Of these, six used YEARS as a screening tool on patients diagnosed with another disease including chronic obstructive pulmonary disease (COPD), sickle cell disease and coronavirus disease. This failed to meet the specified inclusion criteria of 'suspected PE' and is not how YEARS was intended to be applied. Four combined YEARS with additional investigations such as the pulmonary rule-out exclusion criteria rule or C-reactive protein studies and failed to meet the original YEARS protocol. This left 10 papers included for the systematic review [supplementary material Figure 1].

Characteristics of Included Studies

In total, 13,993 participants were included across 10 studies[9,29-37] with no one study making up more than 25% of the total review cohort. Participants were recruited internationally across 11 countries by way of 39 different hospitals. In total, 1.4% of the participants were lost of follow up in two studies.[9,31] Most participants were recruited within the ED.

The incidence of PE varied significantly between studies with an average incidence of 17.2% (SD ±9.9%). Only one study[35] included participants aged 16-17 years old compared to all other studies who opted for 18 years or older. Conversely, one study[29] excluded all participants under the age of 50 years. The key characteristics of the included studies is presented in Table 1.

Methodological Quality

All studies[9,29-37] received critical appraisal with the average result equalling 8.5/10. No studies were at high risk for bias and only one study[35] appearing to be at moderate risk of bias: 7/10. A table with critical appraisal of the included studies is presented in Table 2. A description of how individual studies were scored for each of the 10 questions within the JBI critical appraisal tool is discussed below.

All studies[9,29-37] incorporated consecutive enrolment of participants, avoided a case-control design, used the original YEARS and D-dimer decision threshold, interpreted the reference

test without knowledge of the index test and allowed for a suitable time between index and reference tests (questions 1, 2, 5, 7, and 8). One study[32] allowed for using venous compression ultrasonography when investigating the presence of a DVT. However, this did not alter the original YEARS algorithm hence this study was scored favourably on question 5.

In question 3, one study[29] implemented inappropriate exclusions by omitting all participants under 50-years-old. In question 4, three studies[9,30,31] interpreted the index test without knowledge of the reference test. One study [9] did not blind clinicians to the D-dimer result before participants had YEARS applied to them. This posed a risk of bias as clinicians knew the result of YEARS before recruitment into the study.[25] Nevertheless, as the D-dimer is separate to the reference test, this study[9] was scored favourably (yes) for question 4.

One study[35] was identified as having the potential for missing PE upon review of their stated reference test (question 6). This study retrospectively reviewed D-dimers ordered for suspected PE and utilised a three-month follow up for patients who did not receive AI at their initial visit. Three-month follow up was performed by reviewing for representation to the same ED or further AI ordered. This sub-group made up 73.7% of the cohort. This follow up was considered at higher risk of missing PE as direct patient follow-up was not performed. Additionally, representation to an alternative ED in the area, for which several were available, was not discussed.

In question 9, six of the studies[29,30,33,34,36,37] uniformly received the same reference test as they were retrospective chart reviews of CTPA scans ordered for suspected PE. One study[31] had a significant number of participants lost to follow up who did not undergo AI upon the index visit – 11% of total cohort (question 10). An additional study[9] also documented participants lost to follow up; however, this number was minuscule compared to the cohort size (0.1%) and thus was scored positively.

Review Findings

Upon review, three studies[9,30,32] required recalculation of their data to align with the protocol of this review (see Table 1). In total, an incidence of 25 ISPE were identified, four of which were negative via the YEARS algorithm. Further sub-group analysis was not possible due to insufficient data. However, characteristics documented were malignancy, heart failure, history of VTE, syncope, lower respiratory tract infections (including corona virus disease-

19), asthma, hormone replacement therapy and Chronic Obstructive Pulmonary Disease (COPD).

Heterogeneity was observed in relation to the two different reference tests used for diagnosis. These being either a mix of AI or three-month follow-up or unilateral use of AI. Both strategies were deemed adequate to detect missed PE. Two studies [9, 31] prospectively utilised a mix of AI or three-month follow-up depending on the result of YEARS, which, despite causing heterogeneity held high value as it produced data within the live clinical environment.[25,38]

Sub-group analysis was performed to compare prospective vs. retrospective studies (two groups) and it was found that there was little evidence of heterogeneity across the study designs. This is suggested by the highly overlapped pooled 95% confidence intervals regarding sensitivity and specificity between the two groups (supplementary material Figure 2). The difference in sensitivity was small whereas greater deviation was observed for specificity between different study types. Pooled outcomes observed upon sub-group analysis of the efficacy of YEARS in terms of the risk ratio of advanced imaging utilisation was very similar between the two groups (supplementary material Figure 3). Again, little or no evidence of heterogeneity by study design was observed.

Given the lack of heterogeneity across the study designs, meta-analysis based on all studies was presented in the main body of our paper. Figure 2 shows a forest plot demonstrating the meta-analysis results of sensitivity and specificity.[26] Figure 3 shows the summary ROC plot with effect-analysis.[26] The overall outcome metrics as per the first primary objective were calculated: sensitivity = 96% (95% CI 93-98%) and specificity = 50% (95% CI 33-67%). The sensitivity calculated via meta-analysis held a reassuringly narrow confidence interval suggesting good between-study reproducibility for this metric. This was not the case for the specificity which held a wide confidence interval and was inconsistent. This is shown within the summary ROS plot where the prediction region suggests significant heterogeneity between studies despite an identical decision-threshold being universally applied. Further pooled statistics were calculated: positive and negative predictive values = 29% and 99% and positive and negative likelihood ratios = 2.35 and 0.06.[39]

Six categories of reference tests were identified to compare rates of AI utilisation: Dichotomized WELLS (D-WELLS), altered D-WELLS, three-tier WELLS, age-adjusted

three-tier WELLS, age-adjusted D-WELLS and clinical gestalt. Five studies[9,29,33,35,36] utilised the D-WELLS though one study,[33] despite inferring AI was reduced, did not supply statistical data for this. The author was contacted however we were unable to resolve this query.[33] The most commonly used reference test was D-WELLS whereas the one which consistently faired the strongest against YEARS was age-adjusted D-WELLS.

Published online letters[40,41] reported that the D-WELLS score used as a reference test in one study[30] considered a positive result only if both a score greater than four in addition to a D-dimer level above 500 ng/ml was present. This deviates from any known version of WELLS and would theoretically produce a lower rate of AI utilisation via the 'threshold effect'.[42] A published response to this letter from the authors [41] confirmed they do not support the use of this altered D-WELLS for use in clinical practice. Because of this, data regarding AI utilisation from this study[30] was not used for meta-analysis. In the case of both prospective studies,[9,31] the reference test was retrospectively applied to the same sample population.

Figure 4 shows the risk ratio of AI being required between YEARS and the reference tests available.[26] The combined risk ratio of AI utilisation attributed to the use of YEARS was 0.78 (95% CI 0.67-0.90). This indicates YEARS decreased the risk ratio of AI being required by 22%. The mean reduction of AI utilisation without effect analysis was 11%. Only one study demonstrated a minimal increase in AI utility. As demonstrated the results between studies were varied despite the overall gross reduction of scans which is signified by the relatively wide confidence interval seen upon meta-analysis. Despite this, the confidence interval of the combined data lay outside of the null effect indicating statistical significance.

Discussion

This systematic review evaluated the diagnostic test accuracy of the YEARS algorithm on nearly 14,000 patients. All participants were recruited using a probability sampling strategy via way of 48 different sampling events internationally (including sites used more than once within a different time period). Malignancy, respiratory or cardiac disease, respiratory tract infections, previous VTE, syncope and hormone replacement therapy – these were among the diverse cohort recruited and represent common challenges when investigating PE due to their increased risk of VTE and/or similar clinical presentations.[1]

Upon review of the YEARS algorithm, the combined sensitivity and specificity was demonstrated to be 96% and 50%. The confidence intervals shown in the forest plot suggests

the sensitivity to be largely consistent between studies. Several risks for potential bias were noted within studies risking over representation of the sensitivity and under representation of the specificity. This included a failure to blind clinicians to D-dimer levels in one study[9] and seven studies[29,30,32-34,36,37] being retrospective chart reviews of CTPA requests. It is unknown to what degree this review was affected by these variables, if at all.

In this review, the diagnostic test accuracy of YEARS has been shown to be effective for use in the clinical environment for safely excluding disease in suspected PE. In fact, if the missed ISPE were excluded from the false negatives, for which emerging evidence may encourage, the miss rate would be even lower than what was demonstrated in this review.[43,44] When analysing the ability to correctly detect disease on the other hand the specificity was largely inconsistent and low. This was similar to the original YEARS study which found the specificity to be only 5% more than this review.[9] Despite this it can be reasonably propositioned that the ability to avoid missing true PE is more valuable than the specificity. The fear of missing PE has been acknowledged, at least in part, to be one of the driving factors of over utilising AI and the avoidance of using clinical decision rules and algorithms.[7,45] One of the prospective studies [31] demonstrated a large proportion of patients where AI was requested against the YEARS protocol. This highlights the presence of mistrust felt by clinicians during clinical use. In relation to this, YEARS did hold a reassuringly high sensitivity and low negative likelihood ratio of 0.06. In fact, the rate of missed PE within the combined cohort was only 0.5%. This falls well short of the generally accepted miss rate for PE of 2% indicating YEARS is likely safe for patient use when considering the risk of missed PE.[46]

In combination to this proposition of a low miss rate, YEARS must also reduce unnecessary AI ordering. In this regard YEARS also appeared to hold value as it demonstrated a decrease of 22 percentage points in the risk ratio. This reduction is statistically significant. These results suggest that YEARS is effective at reducing AI utilisation compared to several different forms of alternative PE algorithms. As is demonstrated in the current literature of PE, over-investigation with AI causes increased risk to both the patient and health care system.[6,7,12,13]

No significant selection bias was observed within the participant exclusion criteria listed across all studies. Common exclusions were the presence of YEARS exclusion criteria such as pregnancy, incomplete participant data, recent use of anticoagulants or a life expectancy

less than three months. A concession to this, though minimal in our opinion, was the exclusion of participants aged 50 years or less in one small study.[29] Two sub-groups of patients appeared at risk of falling below the acceptable level of reference testing for PE and made up 12.6% of the total cohort. This was either participants lost during three-month follow up or participants who did not receive AI within one study [35] due to the concerns discussed during critical appraisal.

Regarding prospective vs. retrospective studies, retrospective analysis is often chosen in studies of diagnostic test accuracy due to data being readily available.[25] This can present risks for error when implementing a protocol compared to prospective implementation.[25] For instance, one study[33] decided on whether PE was the most likely diagnosis retrospectively from chart review, depending on whether the patient had a known disease which would explain breathlessness (e.g. COPD). In practice however, the clinical acumen needed for this decision is more complex. In spite of this, sub-group meta-analysis by study design demonstrated minimal differences regarding outcomes.

Another point for consideration were studies which conducted sampling via retrospective data of CTPA ordered for suspected PE. Such studies may have recruited a proportionally higher cohort of individuals who were at high risk for PE compared to the 'typical' patient with suspected PE. To elaborate, it could be surmised that patients who received CTPA, ordered according to the local protocols, were more likely to have PE compared to patients who had PE excluded without CTPA (thus were not recruited). This could risk the results overestimating sensitivity and under estimating specificity.[25,38]

Interestingly, out of the nine studies [9,29-31,33-37] which included comparative data of YEARS versus an alternative algorithm, seven[9,29,31,33-36] indicated the YEARS algorithm did not produce the lowest rate of missed PE. It was not in the scope of this review to compare the diagnostic accuracy of YEARS against alternative algorithms; therefore, no comment can be made on the superiority or inferiority of YEARS concerning the diagnostic test accuracy within this review.

This review has several limitations. Studies published in languages other than English were excluded. Furthermore, grey literature was not included.[47] A single researcher conducted most of the data extraction and critical appraisal. To mitigate this risk, two authors were consulted throughout the process and 20% of the included studies received calibration

exercises of critical appraisal and data extraction to moderate against error and discuss discrepancies.[48]

Conclusion

This review aimed to evaluate the diagnostic test accuracy YEARS when assessing patients presenting with suspected PE. This review concluded that the YEARS algorithm holds a sufficiently high sensitivity to avoid missing true PE. The specificity suggests YEARS has poor accuracy at detecting true PE (without AI). However, despite the relatively poor specificity, the use of AI was reduced compared to other reference tests analysed. It was demonstrated that the studies synthesised included a wide range of ages, demographics, and genders, with variable medical histories and clinical presentations, in varied clinical settings. This suggests that the results from this review can be applied to a wide range of patient demographics seen in clinical practice. Further research on the implementation of YEARS prospectively is needed to accurately demonstrate its outcomes on patient care during live clinical use. As was discussed, the limitations of this review predominantly stemmed from the use of retrospective study methodology. The future of investigating patients presenting with suspected PE remains a common dilemma for clinicians. The YEARS algorithm has been shown to constitute a possible means of safely managing this patient demographic.

Authors' Contribution

SRTH conceptualized the study. SRTH and HDR designed the methodology and validation. All authors were involved in the formal analysis, investigation, visualization and drafting of the manuscript. SRTH handled the project administration. HDR supervised the work. All authors approved the final version of the manuscript.

References

1. Wendelboe AM, Raskob GE. Global Burden of Thrombosis: Epidemiologic Aspects. Circulation Res. 2016;118(9):1340–47. https://doi.org/10.1161/CIRCRESAHA.115.306841.
2. Keller K, Hobohm L, Ebner M, Kresoja K-P, Munzel T, Konstantinides SV, et al. Trends in thrombolytic treatment and outcomes of acute pulmonary embolism in Germany. Eur Heart J. 2020;41(4):522–29. https://doi.org/10.1093/eurheartj/ehz236.
3. Moores LK. Are we overtreating isolated subsegmental pulmonary embolism?: first do no harm. JAMA Intern Med. 2018;178(9):1274–75. https://doi.org/10.1001/jamainternmed.2018.2970.

4. Bariteau A, Stewart LK, Emmett TW, Kline JA. Systematic Review and Meta-analysis of Outcomes of Patients With Subsegmental Pulmonary Embolism With and Without Anticoagulation Treatment. Acad Emerg Med. 2018;25(7):828–35. https://doi.org/10.1111/acem.13399.
5. Van Der Hulle T, Kooiman J, Den Exter PL, Dekkers OM, Klok FA, Huisman MV. Effectiveness and safety of novel oral anticoagulants as compared with vitamin K antagonists in the treatment of acute symptomatic venous thromboembolism: a systematic review and meta-analysis. J Thromb Haemost. 2014;12(3):320–28. https://doi.org/10.1111/jth.12485.
6. Dalen JE, Alpert JS. Diagnosis and Treatment of Pulmonary Embolism: What Have We Learned in the Last 50 Years? Am J Med. 2020;133(4):404–06. https://doi.org/10.1016/j.amjmed.2019.08.049.
7. Wang RC, Miglioretti DL, Marlow EC, Kwan ML, Theis MK, Bowles EJA, et al. Trends in Imaging for Suspected Pulmonary Embolism Across US Health Care Systems, 2004 to 2016. JAMA Netw Open. 2020;3(11):e2026930. https://doi.org/10.1001/jamanetworkopen.2020.26930.
8. Özsu SS, Durmuş ZG, Coşkuner MB, Bülbül Y, Öztuna F, Özlü T. Does the incidence and mortality of pulmonary thromboembolism change over the years? Turk Thorac J. 2017;18(3):78–81. https://doi.org/10.5152/TurkThoracJ.2017.16050.
9. Van Der Hulle T, Cheung WY, Kooij S, Beenen LFM, Bemmel Tv, Van Es J, et al. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. Lancet. 2017;390(10091):289–97. https://doi.org/10.1016/S0140-6736(17)30885-1.
10. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, et al. 2019 European Society of Cardiology Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology. Eur Heart J. 2020;41(4):543–603. https://doi.org/10.1093/eurheartj/ehz405
11. The Royal College of Radiologists. iRefer: Making the best use of clinical radiology. 7th ed. London, UK: The Royal College of Radiologists; 2012.
12. Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. JAMA. 2007;298(3):317–23. https://doi.org/10.1001/jama.298.3.317.

13. Rose Jr TA, Choi JW. Intravenous imaging contrast media complications: the basics that every clinician needs to know. Am J Med. 2015;128(9):943–49. https://doi.org/10.1016/j.amjmed.2015.02.018.
14. Phillips JJ, Straiton J, Staff RT. Planar and SPECT ventilation/perfusion imaging and computed tomography for the diagnosis of pulmonary embolism: A systematic review and meta-analysis of the literature, and cost and dose comparison. Eur J Radiol. 2015;84(7):1392–400. https://doi.org/10.1016/j.ejrad.2015.03.013.
15. Waxman AD, Bajc M, Brown M, Fahey FH, Freeman LM, Haramati LB, et al. Appropriate Use Criteria for Ventilation–Perfusion Imaging in Pulmonary Embolism: Summary and Excerpts. J Nucl Med. 2017;58(5):13N–15N.
16. Stals MAM, Takada T, Kraaijpoel N, van Es N, Buller HR, Courtney DM, et al. Safety and Efficiency of Diagnostic Strategies for Ruling Out Pulmonary Embolism in Clinically Relevant Patient Subgroups: A Systematic Review and Individual-Patient Data Meta-analysis. Ann Intern Med. 2022;175(2):244–55. https://doi.org/10.7326/M21-2625.
17. Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. JBI; 2020.
18. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349:g7647. https://doi.org/10.1136/bmj.g7647.
19. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies. JAMA. 2018;319(4):388–96. https://doi.org/10.1001/jama.2017.19163.
20. Microsoft Corporation. Microsoft Excel [computer program]. Version 16.0.12527.22079. Microsoft Corporation; 2021.
21. Rayyan - a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4.
22. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology. 2015;277(3):617–28. https://doi.org/10.1148/radiol.2015151516.

24. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890. https://doi.org/10.1136/bmj.l6890.
25. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess. 2005;9(12). https://doi.org/10.3310/hta9120.
26. Nyaga VN, Arbyn M. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – a tutorial. Arch Public Health. 2022;80(1). https://doi.org/10.1186/s13690-021-00747-5.
27. Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. Stat Methods Med Res. 2012;21(4):409–26. https://doi.org/10.1177/0962280210392008.
28. Fisher D, Harris R, Bradburn M, Deeks J, Harbord R, Altman D, et al. METAN: Stata module for fixed and random effects meta-analysis. Stat Softw Components: Boston College Dept Econ; 2022.
29. Zahid A, Shahzad S, Ganaie M. Optimizing pre-imaging diagnosis of pulmonary embolism: A comparison of YEARS algorithm with original and simplified WELL'S scores. Chest. 2020;157:A396–A. https://doi.org/10.1016/j.chest.2020.05.444.
30. Abdelaal Ahmed Mahmoud M. Alkhatip A, Donnelly M, Snyman L, Conroy P, Hamza MK, Murphy I, et al. YEARS Algorithm Versus Wells Score: Predictive Accuracies in Pulmonary Embolism Based on the Gold Standard CT Pulmonary Angiography. Crit Care Med. 2020;48(5):704–8. https://doi.org/10.1097/CCM.0000000000004271.
31. Kabrhel C, Van Hylckama Vlieg A, Muzikanski A, Singer A, Fermann GJ, Francis S, et al. Multicenter Evaluation of the YEARS Criteria in Emergency Department Patients Evaluated for Pulmonary Embolism. Acad Emerg Med. 2018;25(9):987–94. https://doi.org/10.1111/acem.13417.
32. Eddy M, Robert-Ebadi H, Richardson L, Bellesini M, Verschuren F, Moumneh T, et al. External validation of the YEARS diagnostic algorithm for suspected pulmonary embolism. J Thromb Haemost. 2020;18(12):3289–95. https://doi.org/10.1111/jth.15083.
33. Nagel SN, Steffen IG, Schwartz S, Hamm B, Elgeti T. Age-dependent diagnostic accuracy of clinical scoring systems and D-dimer levels in the diagnosis of pulmonary embolism with computed tomography pulmonary angiography (CTPA). Eur Radiol. 2019;29(9):4563–71. https://doi.org/10.1007/s00330-019-06039-5.
34. Silva BV, Rigueira J, Aguiar Ricardo I, Mendonca C, Alves Da Silva P, Brito J, et al. Best approach in d-dimer algorithm to exclude pulmonary thromboembolism: a

comparative study. Eur Heart J Cardiovasc Imaging. 2021;22(Supplement 1). https://doi.org/10.1093/ehjci/jeaa356.250.
35. McLenachan CJ, Chua O, Chan BSH, Vecellio E, Chiew AL. Comparison of Wells and YEARS clinical decision rules with D-dimer for low-risk pulmonary embolus patients. Intern Med J. 2019;49(6):739–44. https://doi.org/10.1111/imj.14138.
36. García-Gómez LC, Castilla-Guerra L, de la Vega-Sánchez J, Olmo-Montes FJ, Colmenero-Camacho MÁ. Usefulness of predictive models of pulmonary embolism. Med Clin (Barc). 2020;154(9):338–43. https://doi.org/10.1016/j.medcli.2019.06.022.
37. Castro-Sandoval P, Barrois-Gonzalez R, Galindo-Martin MA, Ruiz-Grinspan MS, Rodriguez-Leal CM. Use of predictive scales for pulmonary thromboembolism in an emergency department. Med Clin (Barc). 2022;159(10):483–5. https://doi.org/10.1016/j.medcle.2022.03.017.
38. Stiell IG, Bennett C. Implementation of Clinical Decision Rules in the Emergency Department. Acad Emerg Med. 2007;14(11):955–9. https://doi.org/10.1111/j.1553-2712.2007.tb02372.x.
39. The Cochrane Collaboration. Review Manager [software]. Version 5.4. The Cochrane Collaboration; 2020.
40. Korevaar DA, Cohen JF, van Es J. YEARS Algorithm Versus Wells' Score: Incomplete Reporting Undermines Study Quality Assessment. Crit Care Med. 2020;48(8):e730. https://doi.org/10.1097/ccm.0000000000004369.
41. Abdelaal Ahmed Mahmoud MAA, Conroy P, Khaled Hamza M, Purcell A, Murphy I, Snyman L. The authors reply. Crit Care Med. 2020;48(12):e1378–e9. https://doi.org/10.1097/ccm.0000000000004701.
42. Tufanaru C, Munn Z, Stephenson M, Aromataris E. Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. Int J Evid Based Healthc. 2015;13(3):196–207. https://doi.org/10.1097/XEB.0000000000000065.
43. Carrier M, Righini M, Le Gal G. Symptomatic subsegmental pulmonary embolism: what is the next step? J Thromb Haemost. 2012;10(8):1486–90. https://doi.org/10.1111/j.1538-7836.2012.04804.x.
44. Dobler CC. Overdiagnosis of pulmonary embolism: definition, causes and implications. Breathe. 2019;15(1):46–53. https://doi.org/10.1183/20734735.0339-2018.
45. Chan TM, Mercuri M, Turcotte M, Gardiner E, Sherbino J, de Wit K. Making Decisions in the Era of the Clinical Decision Rule: How Emergency Physicians Use Clinical

Decision Rules. Acad Med. 2020;95(8):1230–37. https://doi.org/10.1097/ACM.0000000000003098.

46. Lessler AL, Isserman JA, Agarwal R, Palevsky HI, Pines JM. Testing low-risk patients for suspected pulmonary embolism: a decision analysis. Ann Emerg Med. 2010;55(4):316–26.e1. https://doi.org/10.1016/j.annemergmed.2009.12.001.
47. Paez A. Gray literature: An important resource in systematic reviews. J Evid Based Med. 2017;10(3):233–40. https://doi.org/10.1111/jebm.12266.
48. Gerrish K, Lathlean J, editors. The Research Process in Nursing. 7th ed. Oxford, UK: Wiley Blackwell; 2015.

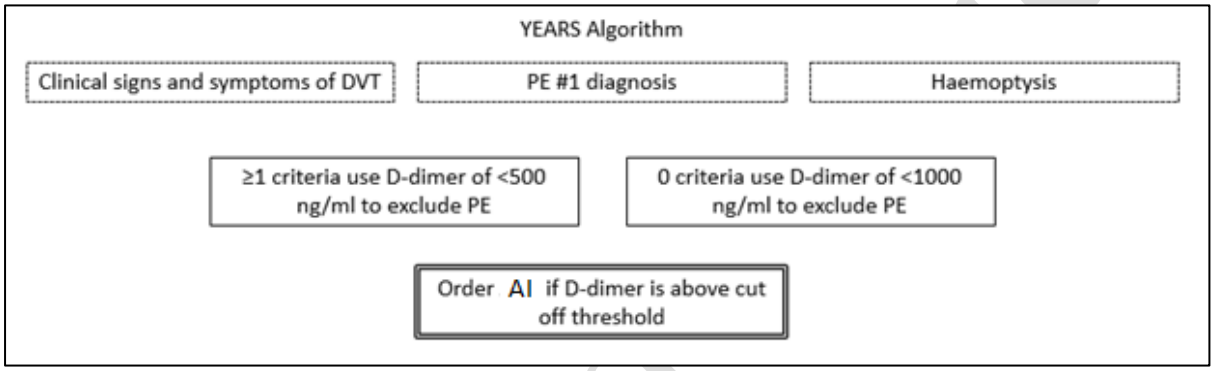

Figure 1: The YEARS Algorithm
Legend: DVT = Deep Vein Thrombosis. PE = Pulmonary Embolism. AI = Advanced Imaging. Adapted from Van Der Hulle T, et al.[9]

Table 1: Data extracted from studies and results of critical appraisal

Other information	1 FN changed to TP (DVT only) AI ordered against protocol (N=40)	Included patients aged 16 years and older	Did not provide comparative data on AI utilisation between YEARS and reference test		Excluded all patients under the age of 50 years	AI ordered against protocol (N=386)	2 FN recalculated to TN (DVT only). 23 ISPE present (21 YEARS +ve and 2 YEARS -ve). Because of this 2 TN changed to 2 FN and 21 FP changed to TP	Statistical error noted p. 706, figure 1: CTPA +ve corrected from 96 to 76		
Female Cohort	62.2%	60.7%	51.2%	UNK	UNK	63.4%	56.9%	59.2%	55.8%	51%
Lost to follow up*	4	N/A	N/A	N/A	N/A	197	N/A	N/A	N/A	N/A
PE Incidence	12.3%	2.2%	25.6%	30.6%	28.9%	4.7%	22.7%	9.8%	20.8%	15.5%
Size	3465	2125	1000	409	353	1789	3314	794	544	200
Recruitment location	12 hospitals in the Netherlands (ED & in/outpatient)	1 ED in Australia	1 ED in Germany	1 ED in Portugal	1 hospital in England (inpatient)	15 ED in the United States	5 EDs in France, Switzerland and Belgium	1 hospital in Ireland (in/outpatient)	1 hospital in Spain (inpatient)	1 ED in Spain
Recruitment strategy	Suspected PE	D-dimer ordered for suspected PE	AI ordered for suspected PE	AI ordered for suspected PE	AI ordered for suspected PE	Suspected PE	AI ordered for suspected PE	AI ordered for suspected PE	AI ordered for suspected PE	AI ordered for suspected PE
Methodology	Prospective	Retrospective	Retrospective	Retrospective	Retrospective	Prospective	Retrospective post-hoc	Retrospective	Retrospective	Retrospective
Study	Van Der Hulle [9]	McLenachan[35]	Nagel [33]	Silva [34]	Zahid [29]	Kabrhel [31]	Eddy [32]	Abdelaal [30]	Garcia-Gomez [36]	Castro-Sandoval [37]

*Legend: FN = False Negative. FP = False Positive. TN = True Negative. TP = True Positive. ED = Emergency Department. N/A = Non Applicable. UNK = Unknown. *Lost to follow up without advanced imaging being performed*

Table 2: Critical appraisal of included studies

	Van Der Hulle [9]	McLenachan [35]	Nagel [33]	Silva [34]	Zahid [29]	Kabrhel [31]	Eddy [32]	Abdelaal [30]	Garcia-Gomez [36]	Castro-Sandoval [37]
1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
4	Y	N	N	N	N	Y	N	Y	N	N
5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
6	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
7	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
8	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
9	N	N	Y	Y	Y	N	N	Y	Y	Y
10	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
T	9	7	9	9	8	8	8	10	9	9

Legend: Y = Yes. N = No. U = Unclear. T = Total score out of 10

1. Was a consecutive or random sample of patients enrolled?
2. Was a case control design avoided?
3. Did the study avoid inappropriate exclusions?
4. Were the index test results interpreted without knowledge of the results of the reference standard?
5. If a threshold was used, was it pre-specified?
6. Is the reference standard likely to correctly classify the target condition?
7. Were the reference standard results interpreted without knowledge of the results of the index test?
8. Was there an appropriate interval between index test and reference standard?
9. Did all patients receive the same reference standard?
10. Were all patients included in the analysis?

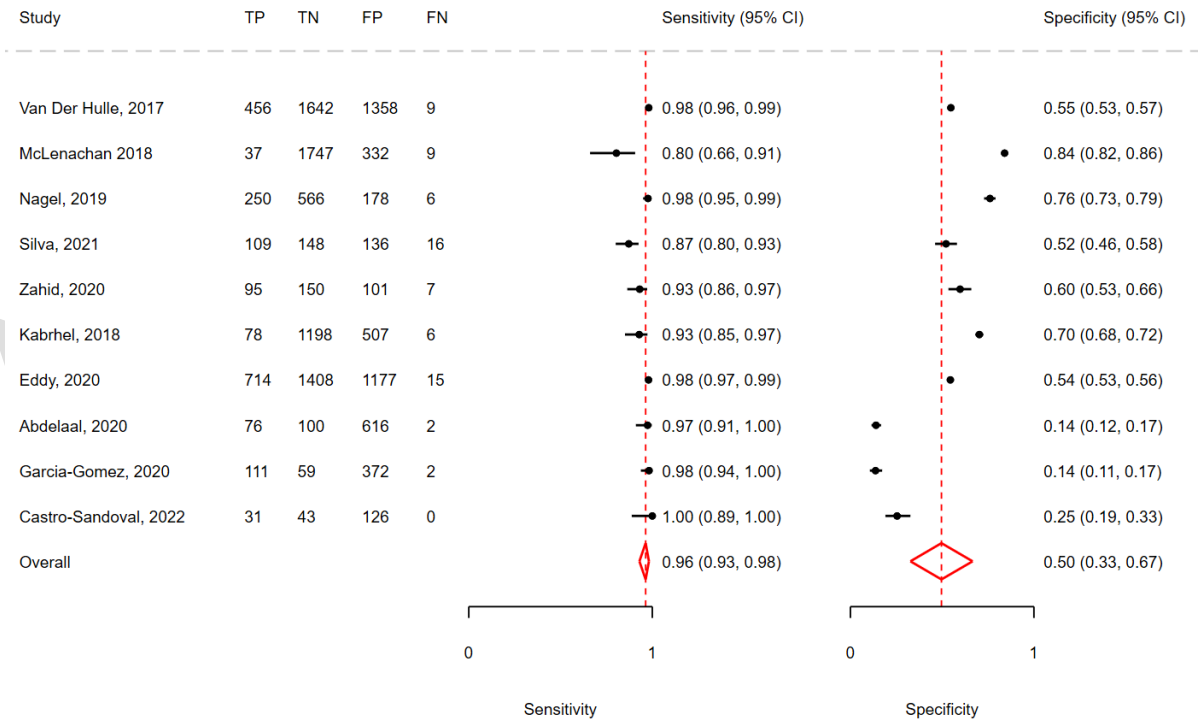

Figure 2: Forest plot of meta-analysis of sensitivity/specificity

Legend: FN = False Negative. FP = False Positive. TN = True Negative. TP = True Positive.

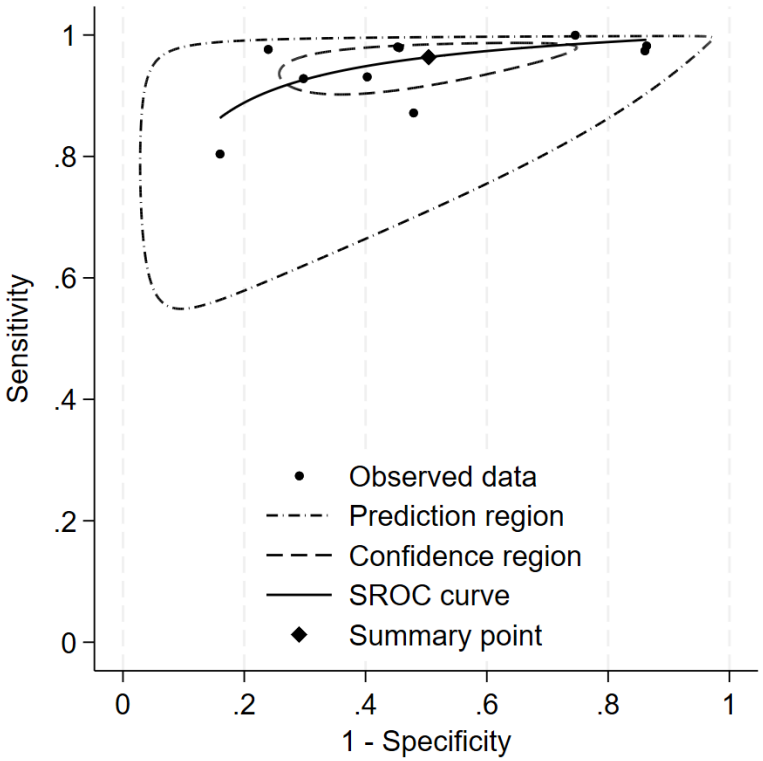

Figure 3: Summary Receiver Operating Characteristics of meta-analysis of diagnostic test accuracy

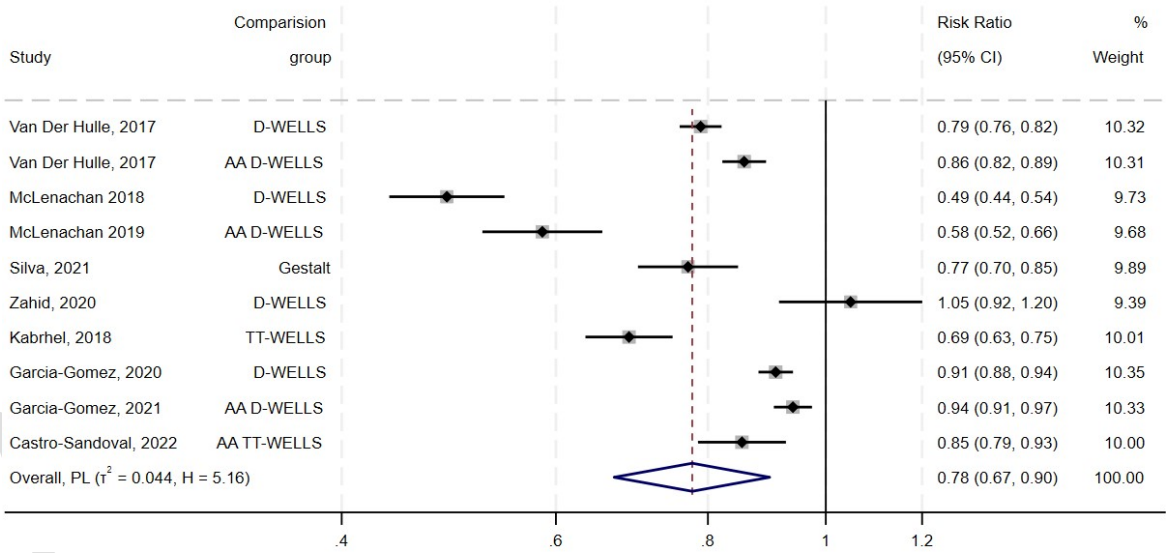

Figure 4: Meta-analysis of risk ratios of AI utilisation of YEARS compared to standard practice

Legend: D-WELLS = Dichotomised WELLS. AA D-WELLS = Age-Adjusted Dichotomised WELLS. TT-WELLS = Three-Tier WELLS. AA TT-WELLS = Age-Adjusted Three-Tier WELLS.