Sultan Qaboos University Med J, November 2024, Vol. 24, Iss. 4, pp. 491–500, Epub. 27 Nov 24
Submitted 14 Aug 23
Revision Req. 25 Oct 23; Revision Recd. 30 Nov 23
Accepted 19 Dec 23
https://doi.org/10.18295/squmj.1.2024.007

REVIEW

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]

ABSTRACT: This systematic review and meta-analysis aimed to assess the diagnostic accuracy of the YEARS algorithm in excluding pulmonary embolism, as well as to compare the utilisation of advanced imaging modalities between the YEARS approach and standard clinical practice. Eligible studies were identified from multiple databases spanning July 2017 to September 2022, following the Joanna Briggs Institute methodology for diagnostic accuracy reviews. A total of 10 studies, involving approximately 14,000 participants, were included in the analysis. The YEARS algorithm demonstrated a sensitivity of 96% (95% confidence interval [CI]: 93–98%) and a specificity of 50% (95% CI: 33–67%). Additonally, the risk ratio for advanced imaging utilisation was 0.78 (95% CI: 67–90), indicating a significant reduction in imaging use. These findings suggest that the YEARS is an effective and safe strategy for 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.

Pulmonary embolism (PE) has an incidence ranging between 39 and 115 per 100,000 individuals and is the third leading cause of cardiovascular mortality worldwide.[1,2] Although PE can present with overt cardiovascular compromise, its clinical manifestations are often non-specific, making diagnosis challenging. As a result, despite its prevalence, PE can be difficult to accurately diagnose, leading to over-investigation, misdiagnosis, prolonged use of emergency department (ED) resources and inappropriate treatment.[3–5] Recent evidence highlights an increasing trend in the utilisation of emergency resources for PE investigation compared to previous decades.[6,7] However, despite this rise in diagnostic efforts, overall mortality rates for PE have not significantly improved.[3,8]

In an effort to diagnose suspected PE while minimising the unnecessary use of advanced imaging (AI), several clinical algorithms have been developed. The YEARS algorithm, first published in *The Lancet* in May 2017, offers a streamlined approach.[9] This algorithm uses a modified version of the WELLS criteria, incorporating 3 key factors: clinical signs and symptoms of deep vein thrombosis (DVT), a likely clinical diagnosis of PE and the presence of haemoptysis. Based on these criteria, patients are categorised as having either a low (i.e. no criteria present) or high (i.e. one or more criteria present) pre-test probability of PE. A D-dimer threshold is then applied, with 500 ng/mL used for low-risk patients and 1,000 ng/mL for high-risk patients. Those within or above the specified D-dimer level requires AI, whereas those below the threshold will have PE excluded on D-dimer alone. Figure 1 presents the YEARS algorithm. The initial study reported a sensitivity of 98%, specificity of 55%, a positive predictive value of 25% and a negative predictive value of 99%.[9]

Computed tomography pulmonary angiography (CTPA) is considered the gold standard for diagnosing PE.[10,11] However, CTPA carries inherent risks, including radiation-induced malignancy, nephrotoxicity and anaphylactic reactions to intravenous contrast agents.[12,13] Furthermore, CT scans are not always readily accessible, and reducing their use could result in cost savings and more efficient use of emergency resources.[14] Although alternative imaging modalities exist that can mitigate some risk, they do not eliminate them entirely.[15] This highlights the need to scrutinise the liberal use of AI in patients with a low or insignificant probability of PE.[11] Accordingly, this review aimed to evaluate the diagnostic test accuracy and utility of the YEARS algorithm in excluding PE.

A preliminary search on CINAHL Plus and Medline identified 1 systematic review related to this topic. The review analysed 4 different algorithms, including the YEARS algorithm, within 4 distinct patient subgroups. Additionally, the review applied a new study protocol retrospectively to cohorts from various studies and compared the YEARS algorithm with 3 other clinical decision rules.[16] Building on this existing evidence, the present review aimed to evaluate the accuracy of the YEARS algorithm in excluding PE and to compare its utilisation of AI with standard clinical practice. This review expanded on

[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*

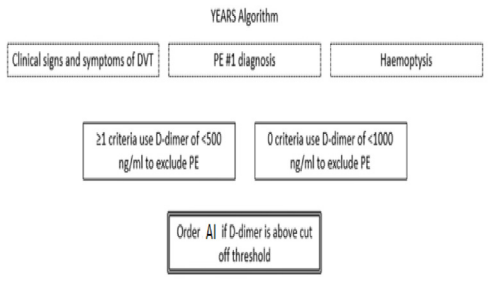

Figure 1: The YEARS algorithm
DVT = deep vein thrombosis; PE = pulmonary embolism; AI = advanced imaging
Adapted from Van Der Hulle et al.[9](2017)

previous research by including studies with varied cohort populations and study design. Furthermore, it employed a broader range of outcome metrics, such as sensitivity, specificity, likelihood ratios and predictive values, rather than focusing solely on missed PE events.

When evaluating an algorithm intended to exclude PE while minimising the use of AI, ensuring patient safety is paramount. This can be demonstrated by estimating both the likelihood of missed PE cases and the extent of AI exposure. The rate of AI utilisation is assessed by comparing it against a reference standard, which typically represents standard clinical practice and may involve alternative algorithms, such as WELLS score, for investigating suspected PE. Thus, the objectives of this review were two-fold (1) to evaluate the accuracy of the YEARS algorithm in excluding PE, as measured by sensitivity, specificity, likelihood ratios and predictive values; and (2) to compare the AI utilisation rate of the YEARS algorithm with standard practice by calculating the associated risk ratio.

Methods

This review followed the Joanna Briggs Institute (JBI) methodology for conducting systematic reviews of diagnostic test accuracy.[17] The review was reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines and the PRISMA extension for diagnostic test accuracy.[18,19]

INCLUSION CRITERIA

The inclusion and exclusion criteria were developed using the population, index test, reference test and diagnosis of interest (PIRD) model.[17,19]

POPULATION

The review included individuals aged 16 years or older, of any ethnicity or gender, with suspected PE, from any geographical location. Studies conducted in EDs, inpatient and outpatient settings were included. Studies involving pregnant participants were excluded.

INDEX TEST

The review used the original YEARS algorithm, with D-dimer thresholds of 500 ng/mL or 1000 ng/mL depending on the YEARS score.

REFERENCE TEST

A reference test was required to confirm the presence or absence of PE, either through AI or prospective application of the YEARS algorithm, with a 3-month follow-up for cases where no AI was performed. This is a common approach in the venous thromboembolism (VTE) literature. Additionally, other PE algorithms were used as reference tests to assess AI utilisation rates.

DIAGNOSIS OF INTEREST

The diagnosis of interest was any thromboembolism within the pulmonary vasculature as detected by AI or autopsy. This review did not differentiate between isolated sub-segmental pulmonary embolism (ISPE) and other forms of PE to maintain homogeneity across the studies. Cases of DVT without PE were not classified as missed PE.

STUDY DESIGN

Randomised control trials, case-control studies, cross-sectional studies and prospective or retrospective cohort studies published in peer-reviewed journals were included. Studies conducting *post-hoc* analyses of literature already included in this review were excluded to prevent duplication of results.

SEARCH STRATEGY

The search strategy was developed in collaboration with a librarian, using the following terms: ('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 conducted in Medline (EBSCO Host, Birmingham, Alabama, USA) to identify keywords and indexed terms. This informed the development of a comprehensive search strategy (listed above) that was adapted for each database (Step 2). The databases searched were CINAHL Plus (EBSCO Host), AMED (EBSCO Host), Medline (EBSCO Host) and EMBASE (Ovid Technologies, Inc., New York, USA).

All studies published in English from July 2017 (the original publication date of the YEARS algorithm) to September 2022 were included.[9] Data was managed using Rayyan™ Reference Manager (Rayyan, Cambridge, Massachusetts, USA) and Microsoft Excel, Version 16.0 (Microsoft Corporation, Redmond, Washington, USA). Two independent reviewers assessed the studies, with disagreements resolved through discussion.

ASSESSMENT OF METHODOLOGICAL QUALITY

The selected studies were critically appraised for methodological quality using the JBI checklist for diagnostic test accuracy studies, which serves as a critical appraisal tool (CAT). This tool is based on 10 'signalling questions' derived from the revised Quality Assessment of Diagnostic Studies (QUADAS-2) tool.[20] The 4 domains used to assess the risk of bias were: (1) patient selection (questions 1–3); (2) index tests (questions 4 & 5); (3) reference standard (questions 6 & 7); and (4) flow and timing (questions 8–10). These domains provided an objective assessment of potential bias in the included studies. All studies were included regardless of their methodological quality. Quality assurance was conducted on 20% of the studies by 2 authors, while the remainder were assessed by a single researcher.

DATA EXTRACTION

Data from the selected studies was extracted by one author using a modified version of the Standards for Reporting Diagnostic Accuracy Studies (STARD) checklist, adapted to align with this review's aims and objectives.[21] Two authors independently performed data extraction on at least 20% of the studies for quality assurance purposes. If the required data was not available, calculations were made based on the information provided. If this was insufficient, the authors of the selected studies were contacted for additional data.

DATA ANALYSIS AND SYNTHESIS

A meta-analysis was conducted following the guidelines outlined below. When meta-analysis was not feasible, data synthesis adhered to the synthesis without meta-analysis (SWiM) reporting guidelines.[22] Meta-analyses for diagnostic test accuracy, focusing on sensitivity and specificity, were conducted according to JBI methodology for systematic reviews of diagnostic test accuracy.[17] The results are presented in a forest plot and a summary receiver operating characterises (ROC) curve.[19,22,23] A hierarchical random-effects logit model was applied, using the 'meta-analysis of diagnostic accuracy' (metadata) package in Stata (StataCorp LLC, Texas, USA).[24]

Additionally, the impact of the YEARS algorithm on reducing the need for imaging scans was compared with other algorithms. Data from these studies were collected, and a meta-analysis of the risk ratio for imaging was conducted, using the profile likelihood method as recommended by Kontopantelis and Reeves,[25] and performed using the 'metan' package in Stata 18 (StataCorp LLC).[26] Results are presented in a forest plot.

Subgroup analyses were also performed to assess heterogeneity between different study designs (prospective versus retrospective) for both outcomes: diagnostic test accuracy and impact on AI utilisation. The same statistical protocol as outlined above was followed, with the results displayed using forest plots. This subgroup analysis was crucial for identifying whether study design significantly contributed to heterogeneity across the included studies.

Results

STUDY INCLUSION

A total of 226 studies were retrieved, and their abstracts were screened. Of these, 25 studies were selected for full-text review, with 15 subsequently excluded for specific reasons. These reasons include English version not available, wrong population (not suspected PE), not original YEARS and background article only. A total of 6 of the excluded studies applied the YEARS algorithm to patients diagnosed with conditions other than PE, such as chronic obstructive pulmonary disease (COPD), sickle cell disease and coronavirus disease. These studies did not meet the inclusion criteria, as they focused on different diseases and were not aligned with the intended use of the YEARS algorithm for suspected PE. Furthermore, 4 studies combined the YEARS algorithm with additional diagnostic criteria, such as the pulmonary embolism rule-out criteria (PERC) or C-reactive protein (CRP), which deviated from the original YEARS protocol. Consequently, 10 studies were included in the systematic review [Supplementary Figure 1].

CHARACTERISTICS OF INCLUDED STUDIES

The 10 selected studies included a total of 13,993 participants, with no single study constituting more than 25% of the overall review cohort.[9,27–35] Participants were recruited internationally from 11 countries, involving 39 different hospitals. A total of 1.4% of participants were lost to follow-up in 2 studies.[9,29] The majority of the studies recruited participants from the EDs.

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

Study	Van Der Hulle *et al.*[9] (2017)	McLenachan *et al.*[33] (2019)	Nagel *et al.*[31] (2019)	Silva *et al.*[32] (2021)	Zahid *et al.*[27] (2020)	Kabrhel *et al.*[29] (2018)	Eddy *et al.*[30] (2020)	Abdelaal et al.[28] (2020)	Garcia-Gomez *et al.*[34] (2020)	Castro-Sandoval *et al.*[35] (2022)
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	3,465	2,125	1,000	409	353	1,789	3,314	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
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		

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*

The incidence of PE varied considerably between the included studies, with an average incidence of 17.2 ± 9.9%. Only 1 study included participants aged 16–17 years,[33] whereas the remaining studies restricted participants to those aged 18 years or older. Conversely, one study excluded participants under the age of 50.[27] The key characteristics of the included studies are summarised in Table 1.

METHODOLOGICAL QUALITY

All studies underwent critical appraisal, with an average score of 8.5/10.[9,27–35] None of the studies were identified as having a high risk of bias. However, one study was rated as having a moderate risk of bias with a score of 7/10.[33] The results of the critical appraisal are detailed in Table 2.

All studies employed consecutive enrolment of participants, avoided a case-control design, used the original YEARS algorithm and D-dimer decision thresholds, interpreted the reference test without prior knowledge of the index test and allowed for a suitable time frame between the index and reference tests (questions 1, 2, 5, 7 and 8).[9,27–35] Venous compression ultrasonography to investigate DVT was used in 1 study, but this did not alter the original YEARS algorithm, and the study was rated favourably on question 5.[30]

For question 3, 1 study excluded participants under the age of 50, which was deemed inappropriate. Regarding question 4, 3 studies interpreted the index test without knowledge of the reference test.[9,28,29] One study did not blind clinicians to the D-dimer results before applying the YEARS algorithm, creating a potential bias as the clinicians knew the results before enrolling participants in the study.[9,23] Despite this, since the D-dimer test is distinct from the reference test, the study received a favourable score for question 4.[9]

A study was identified as having a potential risk of missing PE due to limitations in the reference test used (question 6).[33] This study retrospectively reviewed D-dimer tests ordered for suspected PE and employed a 3-month follow-up for patients who did not receive AI during their initial visit. The follow-up consisted of reviewing subsequent representations to the same ED or further AI orders. This subgroup comprised 73.7% of the cohort. The lack of direct patient follow-up and the failure to address potential representations at alternative EDs in the area, of which several were available, heightened the risk of missed PE.

For question 9, 6 studies uniformly used the same reference test, as they were retrospective chart reviews of CTPA scans ordered for suspected PE.[27,29,31,32,34,35] A substantial portion of 1 study's participants (11% of the total cohort) were lost to follow-up after they did not undergo AI upon the index visit (question 10).[29] Another study also reported a small number of participants lost to follow-up;[9] however, the proportion was negligible (0.1%) relative to the cohort size, and thus the study was scored positively.

REVIEW FINDINGS

Upon review, 3 studies required recalculations to align their data with the protocol of this review [Table 1].[9,28,30] A total of 25 incidences of ISPE were identified, of which 4 were missed by the YEARS algorithm. However, further subgroup analysis was not feasible due to insufficient data. The documented characteristics of the patients included malignancy, heart failure, a history of VTE, syncope, lower respiratory tract infections (including COVID-19), asthma, hormone replacement therapy and COPD.

Heterogeneity was observed across the studies, particularly due to the use of 2 different reference tests for diagnosis. These were either (a) prospective

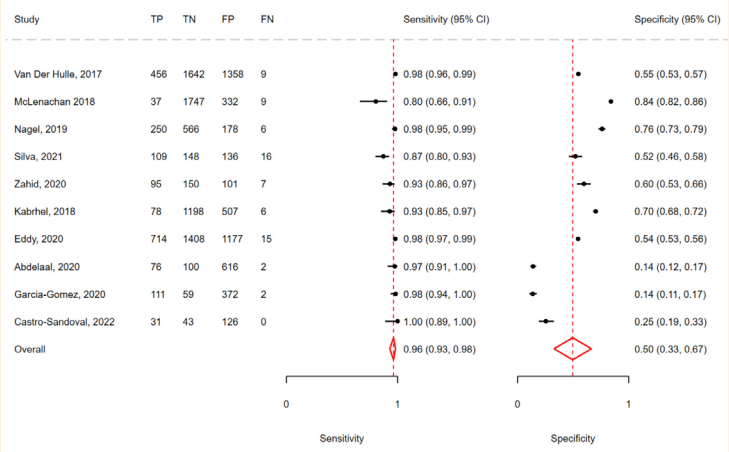

Figure 2: Forest plot of meta-analysis of sensitivity/specificity

FN = false negative; FP = false positive; TN = true negative; TP = true positive

Table 2: Critical appraisal of included studies

	Van Der Hulle *et al.*[9] (2017)	McLenachan *et al.*[33] (2019)	Nagel *et al.*[31] (2019)	Silva *et al.*[32] (2021)	Zahid *et al.*[27] (2020)	Kabrhel *et al.*[29] (2018)	Eddy *et al.*[30] (2020)	Abdelaal *et al.*[28] (2020)	Garcia-Gomez *et al.*[34] (2020)	Castro-Sandoval *et al.*[35] (2022)
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

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 the index test and the reference standard?
9. Did all patients receive the same reference standard?
10. Were all patients included in the analysis?

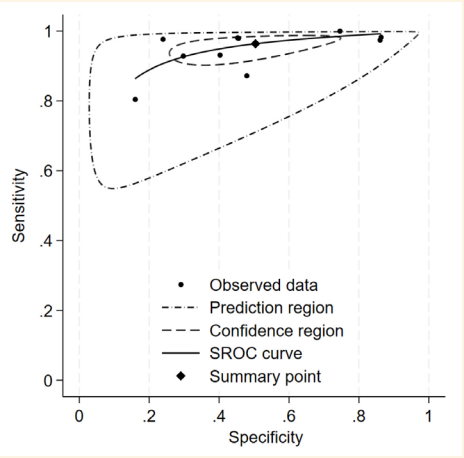

Figure 3: Summary receiver operating characteristics of meta-analysis of diagnostic test accuracy

utilisation of YEARS in which a positive result requires AI and a negative result receives three-month follow-up, or (b) unilateral use of AI (regardless of the result of YEARS). Both strategies were considered adequate to detect missed PE. Two studies prospectively employed a combination of AI or 3-month follow-up depending on the result of the YEARS algorithm. While this caused some heterogeneity, it was valuable as it provided data from a live clinical setting.[23,36]

SUBGROUP ANALYSIS

A subgroup analysis was conducted to compare prospective and retrospective studies. Despite differing study designs, little evidence of heterogeneity was found between the two groups, as indicated by the overlapping pooled 95% CI for both sensitivity and specificity [Supplementary Figure 2]. While sensitivity differed minimally between the groups, a greater variance was noted in specificity. The pooled outcomes from the subgroup analysis of the efficacy of the YEARS algorithm in terms of the risk ratio of AI utilisation were similar between the groups [Supplementary Figure 3]. Once again, no significant heterogeneity by study design was observed.

Given the lack of heterogeneity across study designs, a meta-analysis encompassing all studies was conducted. Figure 2 presents a forest plot demonstrating the meta-analysis results for sensitivity and specificity, while Figure 3 shows the summary ROC plot with effect analysis.[24] The overall outcome metrics for the primary objective were calculated as follows: sensitivity = 96% (95% CI: 93–98%) and specificity = 50% (95% CI: 33–67%). The sensitivity demonstrated a reassuringly narrow CI, indicating good reproducibility across study. However, the specificity was less consistent, reflected by a wide CI. This inconsistency is also depicted in the summary

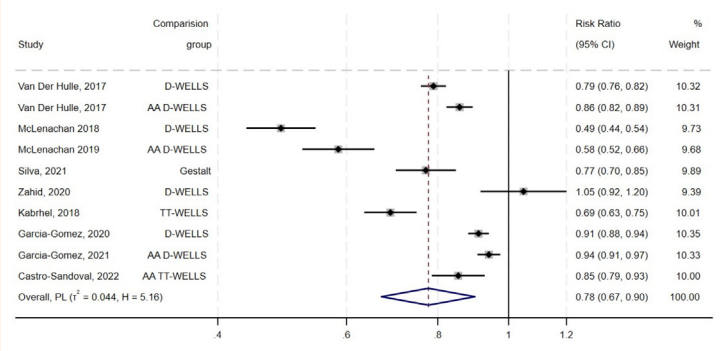

Figure 4: Meta-analysis of risk ratios of AI utilisation of YEARS compared to standard practice

D-WELLS = dichotomised WELLS; AA D-WELLS = age-adjusted dichotomised WELLS; TT-WELLS = three-tier WELLS; AA TT-WELLS = age-adjusted three-tier WELLS

ROC plot, where the prediction region suggests significant heterogeneity between studies, despite the universal application of the same decision threshold. Additional pooled statistics yielded the following: positive predictive value = 29%, negative predictive value = 99%, positive likelihood ratios = 2.35, and negative likelihood ratios= 0.06.[37]

Moreover, 6 categories of reference tests were identified to compare rates of AI utilisation: Dichotomised WELLS (D-WELLS), altered D-WELLS, three-tier WELLS, age-adjusted three-tier WELLS, age-adjusted D-WELLS and clinical gestalt. Overall, 5 studies employed the D-WELLS score,[9,27,31,33,34] though 1 study inferred a reduction in AI but did not provide statistical data to support this claim.[31] Attempts to clarify this with the author were unsuccessful. The D-WELLS was the most commonly used reference test, with age-adjusted D-WELLS consistently outperforming the YEARS algorithm in terms of reducing AI utilisation.

Two published letters raised concerns regarding the use of the D-WELLS score as a reference test in 1 study, where a positive result required both a WELLS score greater than 4 and a D-dimer level >500 ng/mL. [28,38,39] This deviates from any known version of the WELLS score and would likely produce a lower AI utilisation rate via the 'threshold effect'.[40] In response, the authors of the study confirmed that they do not endorse this altered D-WELLS for clinical use.[39] Consequently, the AI utilisation data from that study was excluded for the meta-analysis.[28] For both prospective studies, the reference test was retrospectively applied to the same sample population.[9,29]

Figure 4 illustrates the risk ratio of AI being required between the YEARS algorithm and the various reference tests.[24] The combined risk ratio of AI utilisation with the YEARS algorithm was 0.78 (95% CI: 0.67–0.90), indicating that the YEARS algorithm reduced the risk of requiring AI by 22%. The mean reduction in AI utilisation, without effect analysis, was found to be 11%. Only 1 study demonstrated a minimal increase in AI utilisation.[27] Although there was variability in the results across studies, the overall reduction in scans was evident, as shown by the relatively wide CI in the meta-analysis. The CI 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 in a cohort of nearly 14,000 patients. Participants were recruited using a probability sampling strategy across 48 international sampling events (including repeat visits to some sites during different time periods). The cohort included patients with malignancy, respiratory or cardiac disease, respiratory tract infections, previous VTE, syncope and hormone replacement therapy. These conditions represent significant challenges when investigating PE due to their elevated risk of VTE and similar clinical presentations.[1]

The review found that the combined sensitivity and specificity of the YEARS algorithm were 96% and 50%, respectively. As indicated by the confidence intervals shown in the forest plot, sensitivity was consistent across the studies. However, certain biases may have affected the results, potentially over-representing sensitivity and under-representing specificity. Examples of these biases include the lack of blinding to D-dimer levels in 1 study and the retrospective nature of 7 studies which were chart reviews of CTPA requests.[9,27,28,30–32,34,35] The extent to which these biases influenced the results remains uncertain.

In the clinical environment, the diagnostic accuracy of the YEARS algorithm was shown to be effective for safely excluding PE in suspected cases. Notably, if ISPE were excluded from the false

negatives—something emerging evidence may support—the miss rate of the YEARS algorithm would be even lower than the rate presented in this review.[41,42] Conversely, the specificity of the algorithm was found to be inconsistent and relatively low. This finding aligns with the original YEARS study, which rerported specificity only marginally higher than that of this review.[9] Nevertheless, it can be reasonably argued that avoiding missed true PE cases is more critical than the specificity rate. Indeed, the fear of missing PE is widely recognised as 1 of the driving factors behind the over-utilisation of AI and the reluctance to adopt clinical decision rules and algorithms.[7,43]

One of the prospective studies reported a substantial proportion of patients for whom AI was requested despite the YEARS algorithm indicating otherwise, highlighting clinicians' mistrust of the algorithm during clinical use.[29] Despite this, the YEARS algorithm demonstrated a reassuringly high sensitivity and a low negative likelihood ratio (0.06), suggesting it is highly effective at ruling out PE. The overall rate of missed PE in the combined cohort was only 0.5%, which is significantly below the generally accepted 2% miss rate for PE.[44] This suggests that the YEARS algorithm is likely safe for clinical use, particularly when the focus is on minimising the risk of missed PE.

In addition to the low miss rate, the YEARS algorithm also demonstrated a significant reduction in unnecessary AI orders. Specifically, it was associated with a 22 percentage point decrease in the risk ratio for AI utilisation, a statistically significant finding. This suggests that the YEARS algorithm effectively reduces the reliance on AI compared to several alternative PE diagnostic algorithms. Over-investigation with AI, as highlighted in the literature, increases risks to both patients and healthcare systems.[6,7,12,13]

No significant selection bias was observed regarding the participant exclusion criteria across the studies. Common exclusion criteria included the presence of the YEARS algorithm exclusion criteria, such as pregnancy, incomplete participant data, recent anticoagulant use or a life expectancy of less than 3 months. A minor exception was the exclusion of participants aged 50 years or younger in 1 study.[27] Two sub-groups, representing 12.6% of the total cohort, were identified as potentially falling below the acceptable level of reference testing for PE: participants lost to 3-month follow-up and those who did not receive AI in 1 study due to concerns raised during critical appraisal.[33]

The distinction between prospective and retrospective study designs is noteworthy. Retrospective analyses are commonly employed in diagnostic accuracy studies due to the availability of pre-existing data. However, they may introduce risks of error when implementing a protocol retrospectively compared to prospective application.[23] For example, 1 study determined whether PE was the most likely diagnosis retrospectively by reviewing patient charts for comorbidities (e.g., COPD) that could explain breathlessness. In practice, this clinical decision is more complex and nuanced. Nonethelesss, subgroup meta-analysis by study design indicated minimal differences in diagnostic outcomes between prospective and retrospective studies.

Another consideration is the recruitement of patients through retrospective data from CTPA orders for suspected PE. Such studies may have included a disproportionately high number of individuals at elevated risk for PE compared to the broader population of patients with suspected PE. This could imply that patients undergoing CTPA, in accordance with local protocols, had a higher likelihood of PE than those who did not receive CTPA and were excluded from recruitment. This scenario risks overestimating sensitivity and underestimating specificity in the study results.[23,36]

Among the 9 studies that provided comparative data between the YEARS algorithm and alternative algorithms,[9,27–29,31–35] 7 studies reported that the YEARS algorithm did not achieve the lowest rate of missed PE.[9,27,29,31–34] However, this review did not aim to directly compare the diagnostic accuracy of YEARS with other algorithms. As such, no conclusions can be drawn regarding the superiority or inferiority of the YEARS algorithm in this regard.

Several limitations of this review should be noted. Studies published in languages other than English were excluded and grey literature was not included.[45] Additionally, data extraction and critical appraisal were primarily conducted by a single researcher, which may introduce bias. To mitigate this risk, two additional researchers were consulted during the process, and 20% of the studies underwent calibration exercises for critical appraisal and data extraction to moderate errors and address discrepancies.[46]

Conclusion

This review, which assessed the diagnostic test accuracy of the YEARS algorithm in evaluating patients with suspected PE, concluded that the algorithm possesses sufficiently high sensitivity to reliably exclude true PE cases. However, its specificity was found to be suboptimal in accurately detecting PE without the need for AI. Despite this limitation, the YEARS algorithm successfully reduced AI utilisation

compared to other reference tests. The studies included in this review encompassed a diverse range of patients across various ages, demographics and clinical presentations, suggesting that the results may be generalisable to a broader spectrum of clinical settings.

Further prospective research is necessary to confirm the real-world impact of the YEARS algorithm on patient care during its live clinical application. The retrospective nature of many of the included studies posed limitations, particularly regarding bias and accuracy. Nonetheless, the management of patients with suspected PE remains a significant challenge for clinicians and the YEARS algorithm offers a promising approach to safely navigate this complex clinical scenario.

AUTHORS' CONTRIBUTION

SRTH conceptualised the study. SRTH and HDR designed the methodology and validation. All authors were involved in the formal analysis, investigation, visualisation 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. Circ Res 2016; 118:1340–47. 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; 41:522–9. 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; 178:1274–5. 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; 25:828–35. 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; 12:320–8. 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; 133:404–6. 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: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; 18:78–81. 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; 390:289–97. 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 Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology. Eur Heart J 2020; 41:543–603. 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; 298:317–23. 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;128:943–9. 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; 84:1392–400. 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; 58:13N–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; 175:244–55. https://doi.org/10.7326/M21-2625.
17. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis, 2024. From: https://synthesismanual.jbi.global Accessed: Nov 2023.
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; 319:388–96. https://doi.org/10.1001/jama.2017.19163.
20. 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.
21. 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; 277:617–28. https://doi.org/10.1148/radiol.2015151516.
22. 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.

23. 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; 9:1–113. https://doi.org/10.3310/hta9120.

24. 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: 95. https://doi.org/10.1186/s13690-021-00747-5.

25. 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; 21:409–26. https://doi.org/10.1177/0962280210392008.

26. 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.

27. 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. https://doi.org/10.1016/j.chest.2020.05.444.

28. 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; 48:704–8. https://doi.org/10.1097/CCM.0000000000004271.

29. 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; 25:987–94. https://doi.org/10.1111/acem.13417.

30. 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; 18:3289–95. https://doi.org/10.1111/jth.15083.

31. 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; 29:4563–71. https://doi.org/10.1007/s00330-019-06039-5.

32. 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. https://doi.org/10.1093/ehjci/jeaa356.250.

33. 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; 49:739–44. https://doi.org/10.1111/imj.14138.

34. 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; 154:338–43. https://doi.org/10.1016/j.medcli.2019.06.022.

35. 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; 159:483–5. https://doi.org/10.1016/j.medcle.2022.03.017.

36. Stiell IG, Bennett C. Implementation of clinical decision rules in the emergency department. Acad Emerg Med 2007; 14:955–9. https://doi.org/10.1111/j.1553-2712.2007.tb02372.x.

37. The Cochrane Collaboration. Review manager [software]. Version 5.4. The Cochrane Collaboration. 2020.

38. Korevaar DA, Cohen JF, van Es J. YEARS algorithm versus Wells' score: Incomplete reporting undermines study quality assessment. Crit Care Med 2020; 48:e730. https://doi.org/10.1097/ccm.0000000000004369.

39. Abdelaal Ahmed Mahmoud MAA, Conroy P, Khaled Hamza M, Purcell A, Murphy I, Snyman L. The authors reply. Crit Care Med 2020; 48:e1378–9. https://doi.org/10.1097/ccm.0000000000004701.

40. 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; 13:196–207. https://doi.org/10.1097/XEB.0000000000000065.

41. Carrier M, Righini M, Le Gal G. Symptomatic subsegmental pulmonary embolism: What is the next step? J Thromb Haemost. 2012; 10:1486–90. https://doi.org/10.1111/j.1538-7836.2012.04804.x.

42. Dobler CC. Overdiagnosis of pulmonary embolism: Definition, causes and implications. Breathe 2019; 15:46–53. https://doi.org/10.1183/20734735.0339-2018.

43. 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; 95:1230–37. https://doi.org/10.1097/ACM.0000000000003098.

44. 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; 55:316–26.e1. https://doi.org/10.1016/j.annemergmed.2009.12.001.

45. Paez A. Gray literature: An important resource in systematic reviews. J Evid Based Med 2017; 10:233–40. https://doi.org/10.1111/jebm.12266.

46. Gerrish K, Lathlean J, editors. The Research Process in Nursing. 7th ed. Oxford, UK: Wiley Blackwell, 2015.