Received: 27 January 2023 | Revised: 4 May 2023 | Accepted: 23 May 2023

https://doi.org/10.1016/j.jtha.2023.05.023

ORIGINAL ARTICLE

Accuracy of physicians' intuitive risk estimation in the diagnostic management of pulmonary embolism: an individual patient data meta-analysis

Rosanne van Maanen[1] | Emily S. L. Martens[2] | Toshihiko Takada[3] | Pierre-Marie Roy[4] | Kerstin de Wit[5,6] | Sameer Parpia[5,7] | Noémie Kraaijpoel[8] | Menno V. Huisman[2] | Philip S. Wells[9] | Grégoire Le Gal[9] | Marc Righini[10] | Yonathan Freund[11] | Javier Galipienzo[12] | Nick van Es[13,14] | Jeanet W. Blom[15] | Karel G. M. Moons[16] | Frans H. Rutten[1] | Maarten van Smeden[16] | Frederikus A. Klok[2] | Geert-Jan Geersing[1] | Kim Luijken[16]

[1]Department of General Practice, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

[2]Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, the Netherlands

[3]Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR), Fukushima Medical University, Fukushima, Japan

[4]Department of Emergency Medicine, Angers University Hospital; MitoVasc UMR CNRS 6015 - INSERM 1083, Université Angers, Angers; and French Clinical Research Infrastructure (F-CRIN) Network, INvestigation Network On Venous Thrombo-Embolism (INNOVTE), Saint-Etienne, France

[5]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada

[6]Department of Emergency Medicine, Queen's University, Kingston, Canada

[7]Department of Oncology, McMaster University, Hamilton, Canada

[8]Department of Medicine, Amsterdam University Medical Center, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands

[9]Department of Medicine, University of Ottawa, Ottawa Hospital Research Institute, Ottawa, Canada

[10]Division of Angiology and Hemostasis, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland

[11]Sorbonne University, Emergency Department, Hopital Pitie-Salpêtrière, Assistance Publique—Hopitaux de Paris, Paris, France

[12]Service of Anesthesiology, MD Anderson Cancer Center Madrid, Madrid, Spain

[13]Amsterdam UMC location University of Amsterdam, Department of Vascular Medicine, Meibergdreef 9, Amsterdam, the Netherlands

[14]Amsterdam Cardiovascular Sciences, Pulmonary Hypertension & Thrombosis, Amsterdam, the Netherlands

[15]Department of General Practice, Leiden University Medical Center, Leiden, the Netherlands

[16]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

Correspondence
Rosanne van Maanen, Department of General Practice, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, PO Box 85500, 3508 AB Utrecht, the Netherlands.
Email: r.vanmaanen@umcutrecht.nl

Abstract

Background: In patients clinically suspected of having pulmonary embolism (PE), physicians often rely on intuitive estimation ("gestalt") of PE presence. Although shown to be predictive, gestalt is criticized for its assumed variation across physicians and lack of standardization.

Manuscript handled by: Zhi-Cheng Jing

Final decision: Zhi-Cheng Jing, 23 May 2023

Funding information
Geert-Jan Geersing is supported by a personal Vidi grant from the Dutch Research Council (grant number 91719304). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Objectives: To assess the diagnostic accuracy of gestalt in the diagnosis of PE and gain insight into its possible variation.

Methods: We performed an individual patient data meta-analysis including patients suspected of having PE. The primary outcome was diagnostic accuracy of gestalt for the diagnosis of PE, quantified as risk ratio (RR) between gestalt and PE based on 2-stage random-effect log-binomial meta-analysis regression as well as gestalts' sensitivity and specificity. The variability of these measures was explored across different health care settings, publication period, PE prevalence, patient subgroups (sex, heart failure, chronic lung disease, and items of the Wells score other than gestalt), and age.

Results: We analyzed 20 770 patients suspected of having PE from 16 original studies. The prevalence of PE in patients with and without a positive gestalt was 28.8% vs 9.1%, respectively. The overall RR was 3.02 (95% CI, 2.35-3.87), and the overall sensitivity and specificity were 74% (95% CI, 68%-79%) and 61% (95% CI, 53%-68%), respectively. Although variation was observed across individual studies (I^2, 90.63%), the diagnostic accuracy was consistent across all subgroups and health care settings.

Conclusion: A positive gestalt was associated with a 3-fold increased risk of PE in suspected patients. Although variation was observed across studies, the RR of gestalt was similar across prespecified subgroups and health care settings, exemplifying its diagnostic value for all patients suspected of having PE.

KEYWORDS
diagnosis, pulmonary embolism, venous thromboembolism, venous thrombosis

1 | INTRODUCTION

Pulmonary embolism (PE) is a potentially fatal disease that warrants early detection and treatment [1]. However, the diagnosis of PE is challenging, and delayed diagnosis is common [2]. The classical triad of shortness of breath, pleuritic pain, and hemoptysis is only present in 10% of patients with established PE [3]. The symptoms of shortness of breath and chest pain may also occur in other, often less severe, conditions such as intercostal neuralgia or localized chest myalgia [3,4]. Hemoptysis is more specific but also an uncommon symptom of PE. Considering the potential severity, physicians have a low threshold for additional testing in patients in whom they suspect PE using either D-dimer testing (a plasma biomarker used to rule out thrombosis) or direct referral for computed tomography pulmonary angiography (the reference standard for the diagnosis of PE).

Historically, the decision-making process for this challenging diagnosis was mainly driven by physicians' intuitive judgment called "gestalt", which is usually defined as a clinical impression of whether PE is considered the most likely diagnosis or not. Gestalt estimation has become an important component of clinical decision rules (CDRs) for PE diagnosis [5–7]. Gestalt has, since then, been repeatedly shown to be associated with an increased risk of PE in diagnostic studies [8,9].

Although intuitively appealing, the merit of gestalt in the diagnostic management of patients suspected of having PE has been debated. Several studies have shown that when physicians only used gestalt in the workup of suspected PE, the risk of PE was overestimated compared with the observed prevalence, resulting in decreased overall efficiency of the diagnostic process, with more patients being referred for imaging [10–12]. Another, perhaps an even more important concern is that gestalt estimation is dependent on clinical experience in the diagnosis of PE in daily practice, resulting in variable interobserver reproducibility [13–15]. Thus, the diagnostic accuracy of gestalt in patients suspected of having PE may vary across health care settings due to differences in experience among physicians working in that setting. Furthermore, it is not well studied how gestalt varies across patient characteristics. This knowledge gap needs to be

Essentials

- The diagnostic accuracy of physicians' gestalt estimation of pulmonary embolism (PE) is unclear.
- An individual patient data meta-analysis of 16 studies, including 20 770 patients suspected of having PE, was conducted.
- A positive physicians' gestalt estimation was associated with a 3-fold higher risk of PE.
- The diagnostic accuracy of physicians' gestalt was similar across all subgroups and health care settings.

addressed to understand the diagnostic "behavior" of this subjective item in assessing the risk of PE and for informing physicians in determining the context in which clinical gestalt can be of merit and in which not.

Therefore, this study aimed to quantify the diagnostic accuracy of gestalt in the diagnostic management of patients suspected of having PE across patient, study, and health care characteristics. We performed an ancillary analysis of a large international individual patient data (IPD) dataset, including >35 000 patients suspected of having PE [16].

2 | METHODS

This was an ancillary analysis of a preregistered individual patient data meta-analysis (IPD-MA) (PROSPERO database for systematic reviews number CRD42018089366), for which a protocol has been published [16]. Previous studies using these IPD explored the diagnostic accuracy of existing CDRs for PE across clinically relevant subgroups and health care settings but not of gestalt specifically [17,18]. Ethical approval and informed consent of individual patients were obtained in each included original study. Throughout this IPD-MA, we adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis for Individual Participant Data (PRISMA-IPD) and the Preferred Reporting Items for Diagnostic Test Accuracy (PRISMA-DTA) guidelines on reporting of systematic reviews, including IPD [19,20].

2.1 | Study eligibility, identification, and selection

The systematic search strategy for this IPD-MA, including information sources and the study selection process, was described in detail previously [16]. In short, MEDLINE was searched from January 1, 1995, until November 1, 2021. Studies were eligible if they evaluated diagnostic strategies for PE, had a prospective cohort design, and included patients suspected of having PE with an objectively confirmed diagnosis of venous thromboembolism (VTE) or clinical follow-up of at least 1 month. For the current analysis, we excluded studies that did not assess the (gestalt) variable "PE most likely diagnosis" and studies including only patients with a low clinical pretest probability. Full-text screening was performed independently by 2 couples (G.-J.G. and N.K. as well as F.A.K. and N.v.E.). The corresponding authors of eligible studies were asked to provide deidentified IPD. The risk of bias in the individual studies was independently assessed by 3 pairs of authors (G.-J.G. and T.T., N.v.E. and N.K., and F.A.K. and M.A.M.S.) using the QUADAS-2 tool for the assessment of the risk of bias and applicability of primary diagnostic accuracy studies [21]. Disagreements were solved by discussion within each pair and between pairs. The finally included set consisted of 23 studies, of which 16 were analyzed in the present work (Supplementary Figure S1).

2.2 | Variable measurements

Clinical gestalt was defined as definitions used in the Wells and other diagnostic PE decision rules in the individual studies, ie, "PE is the most likely diagnosis". If PE was considered the most likely diagnosis, the gestalt item was defined as positive, whereas if PE was not considered the most likely diagnosis, the gestalt item was defined as negative. In all studies, physicians were instructed to score the gestalt item before the result of the D-dimer test was known.

The diagnostic accuracy of gestalt was estimated across different patient subgroups, categorized by the following variables: male vs female patients; heart failure (present or absent at presentation with suspected PE); chronic lung disease (defined as chronic obstructive pulmonary disease, asthma, pulmonary fibrosis, or any other chronic lung disease present or absent at presentation with suspected PE); other Wells items (ie, clinical signs/symptoms of deep vein thrombosis (DVT), previous VTE, heart rate >100 beats/min, hemoptysis, immobilization or surgery in the previous 4 weeks, and active malignancy); subgroups of patients without, with 1, or with ≥2 Wells CDR items in addition to the gestalt item; and age on a continuous scale (in years). We also estimated the risk ratio (RR) across 2 publication periods (before 2010 vs in 2010 and later), PE prevalence, and categories of health care settings (hospital or nursing home care, referred secondary or emergency care, primary health care).

In accordance with the original protocol of the IPD-MA [16], we predefined the following health care settings: (i) hospital or nursing home care, (ii) emergency ward or hospital care, (iii) primary health care, and (iv) self-referral emergency care [18]. Five expert panel members (G.-J.G., F.A.K., M.A.M.S., N.K., and N.v.E.) independently categorized each study into 1 of the 4 defined health care settings and discussed disagreements until they reached a consensus. When studies were performed in >1 setting, individual patients were categorized based on the information provided by principal investigators.

2.3 | Missing data

A summary of missing data in each original study is shown in Supplementary Table S2 [5,6,22–35]. Variables were either partially missing (ie, missing in a certain proportion of patients within the study) or systematically missing (ie, completely missing in the study because data were not collected on that particular variable). Systemically missing values were not imputed. Partially missing values were imputed within each study using multiple imputation techniques with all available variables, including the outcome, using the R-package mice [36] unless the variables were missing in >80% of patients [37]. Ten imputed datasets per study were created. The measures of log-RR, logit-sensitivity, and log-specificity were computed in each imputed set and combined using Rubin rules in each study (ie, in the first stage of the meta-analysis; see below) [38].

2.4 | Data analysis

We described the characteristics and prevalence of PE of the included patients stratified by positive vs negative gestalt. Continuous variables were presented as medians with corresponding interquartile ranges (IQRs), and categorical variables were presented as numbers with corresponding percentages. The primary outcome measure of this study was the diagnostic accuracy of the clinical gestalt estimate for PE diagnosis. We quantified the diagnostic accuracy of gestalt as RR, ie, the presence (or risk) of having PE in individuals with a positive vs negative gestalt item, as well as its sensitivity and specificity [39]. We expressed the diagnostic accuracy of gestalt as RR rather than the commonly used diagnostic odds ratio because of noncollapsibility issues of odds ratio [40].

A 2-stage meta-analysis was performed to estimate the overall RR. In the first stage, the RR was estimated in each study using a log-binomial regression model. In the second stage, these estimates were pooled using a separate intercept for each study and a random effect for gestalt using restricted maximum likelihood estimation. The random effect for gestalt allowed studies to differ in the association between gestalt and PE diagnosis because of real differences in RR rather than chance variation only [41]. This resulted in an overall RR and 95% prediction interval (PI) for the association between gestalt and PE diagnosis. Studies without observations in 1 of the cells of the 2×2 table for gestalt vs PE were excluded from the meta-analysis.

To gain an insight into the diagnostic accuracy of gestalt across patient, study, and health care characteristics, we stratified the data into the following subgroups as described above. The 2-stage meta-analysis was performed in each subgroup separately to estimate the subgroup-specific RR of gestalt and a final PE diagnosis. Furthermore, we assessed how the RR varied across age on a continuous scale by fitting a log-binomial model on the stacked imputed data with an interaction between the variable gestalt and age, where age was modeled using a restricted cubic spline with 5 knots (on the percentiles 0.05, 0.275, 0.50, 0.725, and 0.95) for each imputed dataset. Then, the risk of PE across ages 18 to 90 years was predicted from this model under gestalt positive and negative using the stacked imputed data. The RR was computed from the ratio of these predicted risks and plotted with a 95% CI estimated using 500 bootstrap samples.

We plotted the prevalence of PE in each study against the RR of gestalt in each study. We hypothesized that the diagnostic accuracy of gestalt would be related to PE prevalence because it has been shown that the efficiency (true and false negatives) and failure rate (1 − negative predictive value) of diagnostic strategies depend on the prevalence of PE; as PE prevalence increases, the failure rate increases and efficiency decreases [18].

Finally, the overall sensitivity and specificity were calculated for the gestalt item on the final PE diagnosis in the overall sample and in all above-described subgroups using a bivariate generalized linear mixed-effects model on the logit sensitivity and logit specificity of each study [39]. This yielded an estimate and 95% CI for sensitivity and specificity.

All analyses were performed using R, version 4.0.3, particularly using the metafor package [42,43]. The analysis code is publicly available at https://github.com/KLuijken/IPDMA_PE_Gestalt/.

3 | RESULTS

3.1 | Study selection and included patients

The systematic literature search retrieved 3892 unique studies. A total of 23 studies fulfilled the eligibility criteria, and the original IPD were obtained from corresponding authors, resulting in 35 248 unique patients suspected of having PE. Additionally, we excluded 7 more studies: 3 studies that did not assess the variable "PE most likely" [44–46] and 4 studies that selectively included patients with a low clinical pretest probability (ie, studies evaluating the PE rule-out criteria CDR, also known as PERC) [47–50]. Hence, in the current analysis, 16 studies were included, with a total of 20 770 patients (Supplementary Figure S1). The risk of bias in each included study was generally scored as low (Supplementary Figure S2). The characteristics of the included studies are summarized in Table 1 [5,6,22–35]. The prevalence of PE ranged from 7.4% to 40.9% and the percentage of patients with a positive gestalt ranged from 22.1% to 62.1% in the individual studies.

The patient characteristics stratified by the gestalt item are shown in Table 2. The median age was 56.6 years, 60% was women, and 47% had a positive gestalt. Patients with a positive gestalt had concurrent heart failure or chronic lung disease less frequently but had risk factors for PE, namely active malignancy, recent surgery or immobilization, clinical signs of DVT, and/or a history of VTE, more often. The median D-dimer level was higher in patients with a positive gestalt than in patients with a negative gestalt (1001 ng/mL [IQR, 510-2421] vs 582 ng/mL [IQR, 298-1200], respectively).

3.2 | Main outcomes

The overall prevalence of PE was 20%: 29% in the positive gestalt group vs 9% in the negative gestalt group. The point estimates of RR for the association between gestalt and a final PE diagnosis from the individual studies ranged from 1.46 to 7.71 (Figure 1), with a pooled point estimate of 3.02 (95% CI, 2.35-3.87; 95% PI, 1.14-7.94). The estimated RRs for each subgroup are shown in Figure 2. The estimated RR was 3.26 (95% PI, 1.37-7.78) in women and 2.79 (95% PI, 0.93-8.34) in men. Studies with systematic missing data on heart failure (n = 3) [25,30,35] or chronic lung disease (n = 4) [25,30,34,35] were excluded from the subgroup analysis of comorbidities. The estimated RRs were 1.98 (95% PI, 1.42-2.76) and 3.07 (95% PI, 1.06-8.89) for patients with and without heart failure, respectively, and 2.19 (95% PI, 0.62-7.72) and 3.11 (95% PI, 1.03-9,41) for patients with and without chronic lung disease, respectively. The estimated RRs in the 3 different settings were 4.03 (95% PI, 0.09-182.9) for hospital or nursing home care, 2.85 (95% PI, 0.90-8.99) for emergency ward or hospital care, and 3.81 (95% PI, 3.39-4.28) for primary health care.

TABLE 1 Characteristics of included studies of patients suspected of having pulmonary embolism.

Author, year	Country	Health care setting	Number of patients included	PE prevalence (%)	Gestalt+ (%)
Sanson et al. [22], 2000	The Netherlands	Referred secondary care and inpatients	517	30.9	60.6
Perrier et al. [23], 2004	Switzerland	Referred secondary care	965	23.7	29.4
Perrier et al. [24], 2005	Switzerland	Referred secondary care	755	26.1	38.5
Kearon et al. [25], 2006	Canada	Primary health care and inpatients	1123	15.0	48.4
van Belle et al. [26], 2006	The Netherlands	Referred secondary care and inpatients	3296	21.2	61.5
Goekoop et al. [27], 2007	The Netherlands	Referred secondary care	876	12.6	47.4
Righini et al. [28], 2008	Switzerland	Referred secondary care	1692	21.3	45.9
Douma et al. [29], 2011	The Netherlands	Referred secondary care and inpatients	807	23.8	56.5
Galipienzo et al. [30], 2012	Spain	Referred secondary care	240	26.3	22.1
Geersing et al. [6], 2012	The Netherlands	Primary health care	597	12.2	55.6
Schouten et al. [31], 2014	The Netherlands	Primary health care and nursing homes	129	39.8	49.6
Righini et al. [32], 2014	Switzerland	Referred secondary care	3324	19.2	53.1
Mos et al. [33], 2014	The Netherlands	Referred secondary care and inpatients	279	40.9	62.1
Penaloza et al. [34], 2017	France and Belgium	Referred secondary care	705	21.7	45.8
van der Hulle et al. [5], 2017	The Netherlands	Referred secondary care and inpatients	3448	13.7	50.0
Kearon et al. [35], 2019	Canada	Primary health care and inpatients	2017	7.4	21.0

PE, pulmonary embolism.

There were no studies in our selection from the setting "self-referral emergency care". The subgroups defined by the publication period of the study showed a comparable estimated RR of 3.17 (95% PI, 0.76-13.29) for studies published before 2010 and 2.89 (95% PI, 1.15-7.24) for studies published in 2010 and later. The estimated RRs in the subgroups based on the presence or absence of any or more of the other Wells items were comparable, and the point estimates ranged between 2.01 and 3.19.

The plot of the estimated RR for age on a continuous scale shows that the estimated RR decreased with increasing age (Figure 3), albeit

TABLE 2 Clinical characteristics of patients in whom the physician scored pulmonary embolism as the most likely diagnosis (gestalt+) and patients in whom the physician did not score pulmonary embolism as the most likely diagnosis (gestalt−).

Characteristic	Missing proportion (%)[a]	Gestalt+ (n = 9860)	Gestalt− (n = 10 910)	Total (n = 20 770)
Median age, y, (IQR)	0.0	57.0 (42.8-71.0)	56.0 (41.1-70.0)	56.6 (42.0-70.0)
Female, *n* (%)	0.0	5919 (60.0)	6570 (60.2)	12489 (60.1)
Heart failure, *n* (%)	20.0	483 (5.5)	599 (7.0)	1082 (6.2)
Chronic lung disease, *n* (%)	12.8	906 (10.6)	1166 (14.3)	2072 (12.4)
Active malignancy < 6 mo, *n* (%)	0.0	1266 (12.8)	938 (8.6)	2204 (10.6)
Surgery or immobilization < 4 wk, *n* (%)	0.0	1814 (18.4)	1370 (12.6)	3184 (15.3)
Clinical signs of DVT, *n* (%)	0.0	951 (9.6)	587 (5.4)	1538 (7.4)
Hemoptysis, *n* (%)	0.0	477 (4.8)	501 (4.6)	978 (4.7)
History of VTE, *n* (%)	0.0	1653 (16.8)	1244 (11.4)	2897 (13.9)
Heart rate > 100 beats/min, *n* (%)	0.0	2385 (24.2)	2622 (24.0)	5007 (24.1)
Median D-dimer, ng/mL, (IQR)	15.0	1001.0 (510.0-2421.0)	582.0 (298.0-1200.0)	780.0 (354.0-1706.0)
Diagnosis of PE, *n* (%)[b]	0.0	2844 (28.8)	988 (9.1)	3832 (18.4)

DVT, deep vein thrombosis; IQR, interquartile range; PE, pulmonary embolism; VTE, venous thromboembolism.

[a]Systematic missingness that remained after multiple imputation.

[b]After 3 months of follow-up.

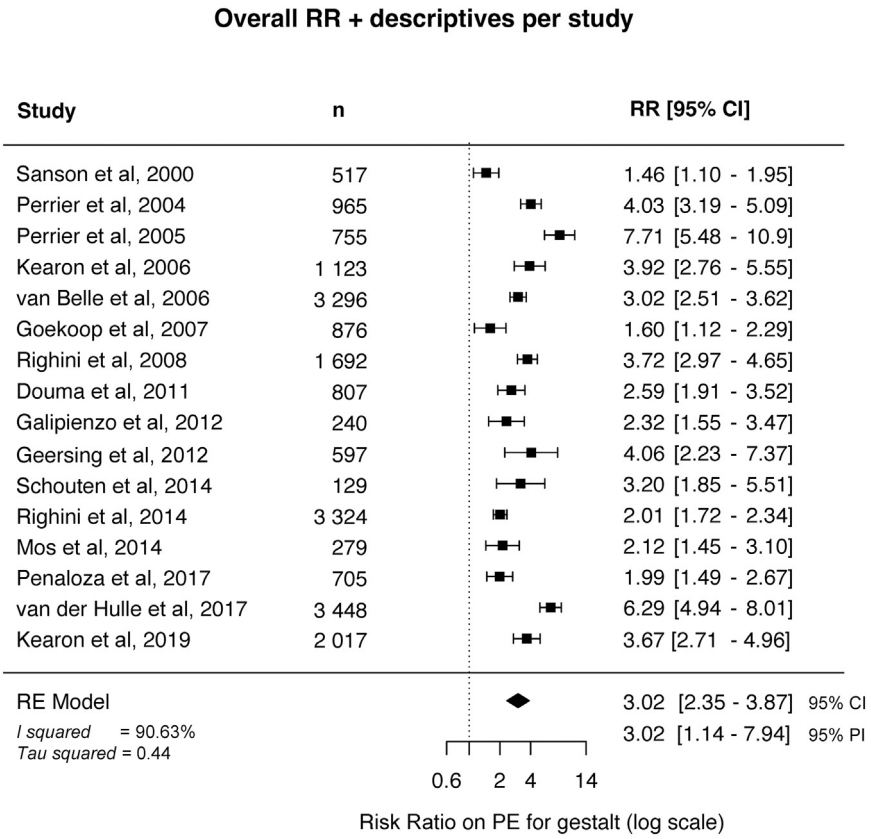

FIGURE 1 Risk ratio of positive gestalt for the presence of pulmonary embolism in individual studies and the pooled estimate. *PI*, prediction interval; *RR*, risk ratio.

with a wide CI in the youngest and oldest patients due to fewer observations in these age groups. The sensitivity and specificity for all subgroups are shown in Supplementary Table S30. The pooled sensitivity of all studies was 74% (95% CI, 68%-79%) and the specificity was 61% (95% CI, 53%-68%), with similar estimates across all evaluated subgroups (Supplementary Table S30). The estimated RR of PE and the gestalt item were plotted against the prevalence of PE in individual studies (Supplementary Figure S110). This did not reveal a clear relation between the diagnostic accuracy of gestalt and the prevalence of PE.

4 | DISCUSSION

In this IPD-MA, which included 20 770 patients suspected of having acute PE, we evaluated the diagnostic accuracy of gestalt in the diagnostic management of PE. Overall, a positive gestalt estimation (ie, a positive score on the item "PE most likely diagnosis") corresponded to a 3-fold higher risk of the presence of PE compared with a negative gestalt estimation. Although there was considerable heterogeneity across individual studies, the overall diagnostic accuracy of gestalt remained remarkably similar across various patient subgroups, health care settings, PE prevalence, and periods of study publication. Only the analysis with age showed that with increasing age, the diagnostic accuracy, on average, decreases, albeit with a wide CI, particularly at higher ages due to fewer observations.

In our extensive subgroup analyses in this large IPD-MA, we could not identify any patient subgroup, clinical setting, or underlying PE prevalence for which gestalt is not beneficial in the diagnostic management of PE. Interestingly, our findings do not support the previous hypothesis that large variability and subjectivity of gestalt would severely hamper its diagnostic accuracy in different health care populations. As an example, much attention was paid to the report of a previous study that the pretest probability of PE increases with clinical experience in managing PE cases [15]. This seems to contradict our finding that the diagnostic accuracy of gestalt did not vary across health care settings, a proxy for physician experience, although of course, it can be debated whether differences in physician experience are fully captured by differences in health care setting. Another previous study showed that with every additional point in the Wells rule, patients had a 1.2-fold increased chance of being assigned the subjective "PE most likely diagnosis" [9]. Indeed, we observed that in patients with an active malignancy, recent surgery or immobilization, clinical signs of DVT, and a history of VTE, the gestalt item is more frequently positively scored. However, the diagnostic accuracy of gestalt, expressed as RR, was comparable across patient subgroups based on the presence of these Wells items when assessed in isolation for each Wells also when patients were stratified by combined presence of other Wells items. The D-dimer values were higher in patients with a positive gestalt than in those with a negative gestalt. However, in some studies—for instance the YEARS study [5]—the D-dimer result

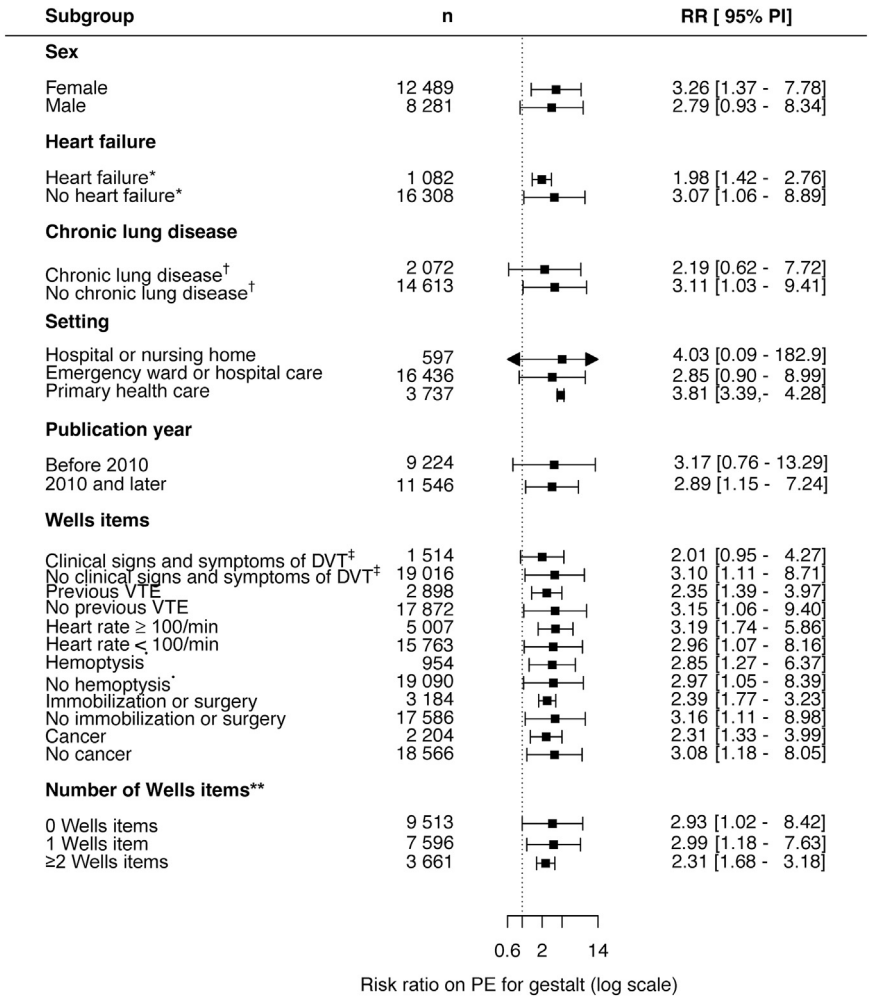

FIGURE 2 Risk ratio of positive gestalt for the presence of pulmonary embolism in predefined subgroups. *3 studies were excluded because of systematic missing and 2 studies with empty cells in the 2 × 2 table in the subgroup with heart failure, †4 studies were excluded because of systematic missing and 1 study with empty cells in the 2 × 2 table in the subgroup with chronic lung disease, ‡1 study was excluded because of empty cells in the 2 × 2 table, •2 studies were excluded because of empty cells in the 2 × 2 table, ¶1 study was excluded because of empty cells in the 2 × 2 table. *DVT*, deep vein thrombosis; *PI*, prediction interval; *RR*, risk ratio; *VTE*, venous thromboembolism.

may already have been known before the scoring of the subjective item "PE most likely diagnosis." This has likely resulted in an "overestimation" of the accuracy of gestalt alone in these studies [51].

4.1 | Strengths and limitations

We performed a comprehensive IPD-MA including data from many individual patients suspected of having PE to study the diagnostic accuracy of gestalt. This allowed us to perform subgroup analyses and provide precise estimates of the diagnostic accuracy of gestalt across different health care settings and patient subgroups. We performed multilevel imputation of missing values and state-of-the-art statistical methods to quantify the diagnostic value of gestalt in patients with suspected PE.

Yet, for full appreciation, several limitations must be discussed. The most important limitation is that the subjective gestalt item was scored in various ways in the individual studies. For instance, in some studies (n = 12), the gestalt estimation was part of the CDR (ie, the Wells rule and YEARS algorithm) and, thus, was scored in the context of these CDRs. On the other end of the spectrum, we included studies evaluating the Geneva rule, in which the gestalt item was scored only for research purposes, thus not being part of a CDR. In studies conducted in the primary care setting, gestalt was always scored before knowing the D-dimer result, whereas likely in some studies conducted in the hospital setting, the gestalt estimate may, to some extent, have been influenced by D-dimer when the result was available before the CDR was assessed [52]. Nevertheless, maybe counterintuitive, the highest RR for gestalt was found in a hospital-based study evaluating the Geneva score (7.71; 95% PI, 5.48,-10.90) [24] and the lowest in a

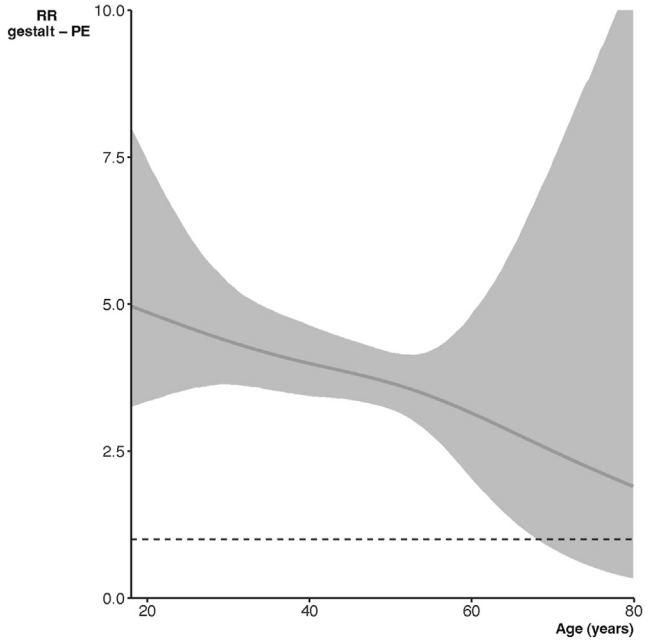

FIGURE 3 Risk ratio of pulmonary embolism for positive vs negative gestalt against age on a continuous scale. *PE*, pulmonary embolism; *RR*, risk ratio.

hospital-based study evaluating the Wells rule (1.46; 95% PI, 1.10-1.95) [22]. Thus, we believe that the overall estimate of a 3-fold higher risk of PE in patients in whom gestalt is scored positively seems to be reasonable and clearly >1.

Another limitation is that the number of patients in the subgroups with heart failure and chronic lung disease was relatively low, and 4 studies did not record these comorbidities. Similarly, for the subgroups with the Wells items "clinical signs/symptoms of DVT" and "hemoptysis," the counts in some studies were low, resulting in convergence issues. Therefore, due to empty cells in these studies in our 2-stage meta-analytical approach, we had to exclude these studies from the specific subgroup analyses.

4.2 | Interpretation of the main findings and the clinical implications

When interpreting our findings, it is important to acknowledge that the focus of this IPD-MA was to explore the variability of the diagnostic accuracy of the gestalt item in the diagnostic management of PE as a stand-alone item in different patient subgroups and health care settings. The goal was not to define whether a CDR should or should not include gestalt estimation. We did not perform such multivariable analyses exploring the incremental diagnostic value of gestalt beyond other CDR items.

From a clinical perspective, we believe that the following conclusions can be drawn: although heterogeneity across individual studies was observed, the diagnostic accuracy of gestalt remained stable, with, on average, a 3-fold increased risk of PE in patients with a positive gestalt across all evaluated patient subgroups and health care settings. Hence, this heterogeneity of gestalt across studies in this IPD-MA did not seem explained by differences in case-mix or health care settings among these individual studies. Although speculative, our analyses suggest that the diagnostic accuracy of the intuitive gestalt item ("gut feeling" of physicians on PE presence) does not substantially vary across risk factors for VTE, sex, age, comorbidity of patients, or health care setting. Rather, it could be related to other factors that are harder to define, such as physicians' expertise and exposure to PE or physicians' clinical impression of the severity of the disease [53]. This is supported by previous work: if physicians experienced a "sense of alarm" in patients with shortness of breath, the odds of having a life-threatening disease increased about 2 folds [54]. Based on our analyses, we might conclude that gestalt estimation or the "sense of alarm" holds its merit, albeit with remaining not fully explained heterogeneity, across all patients with suspected PE, regardless of the health care setting in which they present themselves or the subgroup they belong to. The heterogeneity may be related to the different ways of estimating gestalt in the individual studies. Only the analysis with age showed that the diagnostic accuracy of gestalt might be slowly declining with increasing age. Indeed, the diagnosis of PE can be challenging in elderly patients given the subtle signs and symptoms, presence of other cardiac or pulmonary comorbidities that may mimic PE symptoms, and that frailty may negatively impact the accuracy of diagnostic tests [55,56].

5 | CONCLUSION

A positive gestalt estimation in the diagnostic management of PE predicts, on average, a 3-fold higher risk of PE in suspected patients compared with a negative gestalt estimation. Although heterogeneity was observed across individual studies, the diagnostic accuracy of gestalt remained stable across different subgroups of patients and health care settings. Our study thereby exemplifies the merit of gestalt when assessing a patient suspected of having PE: irrespective of health care setting or subgroup, this subjective variable was indeed associated with a 3-fold increased risk of PE, and this thus calls for acting on it by referring the patient for an appropriate diagnostic workup.

AUTHOR CONTRIBUTIONS

G.-J.G., F.A.K., R.v.M., E.S.L.M., M.v.S., and K.M. conceived and designed the study. G.-J.G., F.A.K., R.v.M., E.S.L.M., F.H.R., T.T., K.G.M.M., M.v.S., and K.L. were involved in analysis and interpretation of the data. G.-J.G., F.A.K., R.v.M., E.S.L.M., F.H.R., and K.L. drafted the article. R.v.M., E.S.L.M., T.T., P.-M.R., K.d.W., S.P., N.K., M.V.H., P.S.W., G.I.G., M.R., Y.F., J.G., N.v.E., J.W.B., K.G.M.M., F.H.R., M.v.S., F.A.K., G.-J.G., and K.L. were involved in critical revision of the article for important intellectual content. R.v.M., E.S.L.M., T.T., P.-M.R., K.d.W., S.P., N.K., M.V.H., P.S.W., G.I.G., M.R., Y.F., J.G., N.v.E., J.W.B., K.G.M.M., F.H.R., M.v.S., F.A.K., G.-J.G., and K.L. were involved in final

approval of the article. Y.F., G.L.G., S.P., M.R., P.-M.R., P.S.W., G.-J.G., and J.G. provided study materials or patients. T.T., K.G.M.M., M.v.S., and K.L. performed the statistical analysis. M.R. and G.-J.G. obtained funding.

DECLARATION OF COMPETING INTERESTS

There are no competing interests to disclose.

TWITTER

Emily S.L. Martens @ESLmartens

REFERENCES

[1] Huisman MV, Barco S, Cannegieter SC, Le Gal G, Konstantinides SV, Reitsma PH, Rodger M, Noordegraaf AV, Klok FA. Pulmonary embolism. *Nat Rev Dis Prim*. 2018;4:18028.

[2] van Maanen R, Trinks-Roerdink EM, Rutten FH, Geersing GJ. A systematic review and meta-analysis of diagnostic delay in pulmonary embolism. *Eur J Gen Pract*. 2022;28:165–72.

[3] Meyer G, Roy PM, Gilberg S, Perrier A. Pulmonary embolism. *BMJ*. 2010;340:974–6.

[4] Erkens PMG, Lucassen WAM, Geersing GJ, van Weert HCPM, Kuijs-Augustijn M, van Heugten M, Rietjens L, ten Cate H, Prins MH, Büller HR, Hoes AW, Moons KGM, Oudega R, Stoffers HEJH. Alternative diagnoses in patients in whom the GP considered the diagnosis of pulmonary embolism. *Fam Pract*. 2014;31:670–7.

[5] van der Hulle T, Cheung WY, Kooij S, Beenen LFM, van Bemmel T, van Es J, Faber LM, Hazelaar GM, Heringhaus C, Hofstee H, Hovens MMC, Kaasjager KAH, van Klink RCJ, Kruip MJHA, Loeffen RF, Mairuhu ATA, Middeldorp S, Nijkeuter M, van der Pol LM, Schol-Gelok S, et al. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. *Lancet*. 2017;390:289–97.

[6] Geersing GJ, Erkens PMG, Lucassen WAM, Büller HR, Ten Cate H, Hoes AW, Moons KGM, Prins MH, Oudega R, Van Weert HCPM, Stoffers HEJH. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in Primary care: prospective cohort study. *BMJ*. 2012;345:1–10.

[7] Le Gal G, Righini M, Roy PM, Sanchez O, Aujesky D, Bounameaux H, Perrier A. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med*. 2006;144:165–71.

[8] Klok FA, Djurabi RK, Nijkeuter M, Huisman MV. Alternative diagnosis other than pulmonary embolism as a subjective variable in the wells clinical decision rule: not so bad after all. *J Thromb Haemost*. 2007;5:1079–80.

[9] Klok FA, Zidane M, Djurabi RK, Nijkeuter M, Huisman MV. The physician's estimation "alternative diagnosis is less likely than pulmonary embolism" in the Wells rule is dependent on the presence of other required items. *Thromb Haemost*. 2008;99:244–5.

[10] Zarabi S, Chan TM, Mercuri M, Kearon C, Turcotte M, Grusko E, Barbic D, Varner C, Bridges E, Houston R, Eagles D, de Wit K. Physician choices in pulmonary embolism testing. *CMAJ*. 2021;193:E38–46.

[11] Kline JA, Stubblefield WB. Clinician gestalt estimate of pretest probability for acute coronary syndrome and pulmonary embolism in patients with chest pain and dyspnea. *Ann Emerg Med*. 2014;63:275–80.

[12] Hendriksen JMT, Lucassen WAM, Erkens PMG, Stoffers HEJH, van Weert HCPM, Büller HR, Hoes AW, Moons KGM, Geersing GJ. Ruling out pulmonary embolism in primary care: comparison of the diagnostic performance of "gestalt" and the wells rule. *Ann Fam Med*. 2016;14:227–34.

[13] Barais M, Morio N, Cuzon Breton A, Barraine P, Calvez A, Stolper E, Van Royen P, Liétard C. "I can't find anything wrong: It must be a pulmonary embolism": diagnosing suspected pulmonary embolism in primary care, a qualitative study. *PLoS One*. 2014;9:1–8.

[14] Rodger MA, Maser E, Stiell I, Howley HEA, Wells PS. The interobserver reliability of pretest probability assessment in patients with suspected pulmonary embolism. *Thromb Res*. 2005;116:101–7.

[15] Kabrhel C, Camargo CA, Goldhaber SZ. Clinical gestalt and the diagnosis of pulmonary embolism: does experience matter? *Chest*. 2005;127:1627–30.

[16] Geersing GJ, Kraaijpoel N, Büller HR, van Doorn S, van Es N, Le Gal G, Huisman MV, Kearon C, Kline JA, Moons KGM, Miniati M, Righini M, Roy PM, van der Wall SJ, Wells PS, Klok FA. Ruling out pulmonary embolism across different subgroups of patients and healthcare settings: protocol for a systematic review and individual patient data meta-analysis (IPDMA). *Diagnostic Progn Res*. 2018;2:1–8.

[17] Stals MAM, Takada T, Kraaijpoel N, van Es N, Büller HR, Courtney DM, Freund Y, Galipienzo J, Le Gal G, Ghanima W, Huisman MV, Kline JA, Moons KGM, Parpia S, Perrier A, Righini M, Robert-Ebadi H, Roy PM, van Smeden M, Wells PS, et al. Safety and efficiency of diagnostic strategies for ruling out pulmonary embolism in clinically relevant patient subgroups. *Ann Intern Med*. 2022;175:244–55.

[18] Geersing GJ, Takada T, Klok FA, Büller HR, Courtney DM, Freund Y, Galipienzo J, Le Gal G, Ghanima W, Kline JA, Huisman MV, Moons KGM, Perrier A, Parpia S, Robert-Ebadi H, Righini M, Roy PM, van Smeden M, Stals MAM, Wells PS, et al. Ruling out pulmonary embolism across different healthcare settings: a systematic review and individual patient data meta-analysis. *PLoS Med*. 2022;19:e1003905.

[19] Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF. Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *JAMA*. 2015;313:1657–65.

[20] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA, Hyde CJ, Korevaar DA, Leeflang MMG, Macaskill P, Reitsma JB, Rodin R, Rutjes AWS, Salameh JP, Stevens A, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies the PRISMA-DTA statement. *JAMA*. 2018;319:388–96.

[21] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, QUADAS-2 Group*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.

[22] Sanson BJ, Lijmer JG, Mac Gillavry MR, Turkstra F, Prins MH, Büller HR. Comparison of a clinical probability estimate and two clinical models in patients with suspected pulmonary embolism. *Thromb Haemost*. 2000;83:199–203.

[23] Perrier A, Roy PM, Aujesky D, Chagnon I, Howarth N, Gourdier AL, Leftheriotis G, Barghouth G, Cornuz J, Hayoz D, Bounameaux H. 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;116:291–9.

[24] Perrier A, Roy PM, Sanchez O, Le Gal G, Meyer G, Gourdier AL, Furber A, Revel MP, Howarth N, Davido A, Bounameaux H. Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med*. 2005;352:1760–8.

[25] Kearon C, Ginsberg JS, Douketis J, Turpie AG, Bates SM, Lee AY, Crowther MA, Weitz JI, Brill-Edwards P, Wells P, Anderson DR, Kovacs MJ, Linkins LA, Julian JA, Bonilla LR, Gent M. An evaluation of D-dimer in the diagnosis of pulmonary embolism: a randomized trial. *Ann Intern Med*. 2006;144:812–21.

[26] van Belle A, Büller HR, Huisman MV, Huisman PM, Kaasjager K, Kamphuisen PW, Kramer MHH, Kruip MJHA, Kwakkel-van Erp JM, Leebeek FWG, Nijkeuter M, Prins MH, Sohne M, Tick LW. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA*. 2006;295:172–9.

[27] Goekoop RJ, Steeghs N, Niessen RWLM, Jonkers GJPM, Dik H, Castel A, Werker-van Gelder L, Vlasveld LT, van Klink RCJ, Planken EV, Huisman MV. Simple and safe exclusion of pulmonary embolism in outpatients using quantitative D-dimer and Wells' simplified decision rule. *Thromb Haemost*. 2007;97:146–50.

[28] Righini M, Le Gal G, Aujesky D, Roy PM, Sanchez O, Verschuren F, Rutschmann O, Nonent M, Cornuz J, Thys F, Le Manach CP, Revel MP, Poletti PA, Meyer G, Mottier D, Perneger T, Bounameaux H, Perrier A. Diagnosis of pulmonary embolism by multidetector CT alone or combined with venous ultrasonography of the leg: a randomised non-inferiority trial. *Lancet*. 2008;371:1343–52.

[29] Douma RA, Mos ICM, Erkens PMG, Nizet TAC, Durian MF, Hovens MM, van Houten AA, Hofstee HMA, Klok FA, ten Cate H, Ullmann EF, Büller HR, Kamphuisen PW, Huisman MV. Performance of 4 clinical decision rules in the diagnostic management of acute pulmonary embolism: a prospective cohort study. *Ann Intern Med*. 2011;154:709–18.

[30] Galipienzo J, Garcia de Tena J, Flores J, Alvarez C, Garcia-Avello A, Arribas I. Effectiveness of a diagnostic algorithm combining clinical probability, D-dimer testing, and computed tomography in patients with suspected pulmonary embolism in an emergency department. *Rom J Intern Med*. 2012;50:195–202.

[31] Schouten HJ, Geersing GJ, Oudega R, Van Delden JJM, Moons KGM, Koek HL. Accuracy of the Wells clinical prediction rule for pulmonary embolism in older ambulatory adults. *J Am Geriatr Soc*. 2014;62:2136–41.

[32] Righini M, Van Es J, Den Exter PL, Roy PM, Verschuren F, Ghuysen A, Rutschmann OT, Sanchez O, Jaffrelot M, Trinh-Duc A, Le Gall C, Moustafa F, Principe A, Van Houten AA, Ten Wolde M, Douma RA, Hazelaar G, Erkens PMG, Van Kralingen KW, Grootenboers MJJH, et al. Age-adjusted D-dimer cutoff levels to rule out pulmonary embolism: the ADJUST-PE study. *JAMA*. 2014;311:1117–24.

[33] Mos ICM, Douma RA, Erkens PMG, Kruip MJHA, Hovens MM, van Houten AA, Hofstee HMA, Kooiman J, Klok FA, Büller HR, Kamphuisen PW, Huisman MV. Diagnostic outcome management study in patients with clinically suspected recurrent acute pulmonary embolism with a structured algorithm. *Thromb Res*. 2014;133:1039–44.

[34] Penaloza A, Soulié C, Moumneh T, Delmez Q, Ghuysen A, El Kouri D, Brice C, Marjanovic NS, Bouget J, Moustafa F, Trinh-Duc A, Le Gall C, Imsaad L, Chrétien JM, Gable B, Girard P, Sanchez O, Schmidt J, Le Gal G, Meyer G, et al. Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study. *Lancet Haematol*. 2017;4:e615–21.

[35] Kearon C, de Wit K, Parpia S, Schulman S, Afilalo M, Hirsch A, Spencer FA, Sharma S, D'Aragon F, Deshaies JF, Le Gal G, Lazo-Langner A, Wu C, Rudd-Scott L, Bates SM, Julian JA. Diagnosis of pulmonary embolism with D-dimer adjusted to clinical probability. *N Engl J Med*. 2019;381:2125–34.

[36] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67.

[37] Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, Moons KGM. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol*. 2010;63:721–7.

[38] Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57.

[39] Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982–90.

[40] Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res England*. 2016;25:1925–37.

[41] Debray TPA, Moons KGM, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RHH, Reitsma JB. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods*. 2015;6:293–309.

[42] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36:1–48.

[43] R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Core Team; 2016.

[44] 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;39:144–52.

[45] 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;161:92–7.

[46] Ghanima W, Almaas V, Aballi S, Dörje C, Nielssen BE, Holmen LO, Almaas R, Abdelnoor M, Sandset PM. Management of suspected pulmonary embolism (PE) by D-dimer and multi-slice computed tomography in outpatients: an outcome study. *J Thromb Haemost*. 2005;3:1926–32.

[47] 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;129:1417–23.

[48] Kline JA, Courtney DM, Kabrhel C, Moore CL, Smithline HA, Plewa MC, Richman PB, O'Neil BJ, Nordenholz K. Prospective multicenter evaluation of the pulmonary embolism rule-out criteria. *J Thromb Haemost*. 2008;6:772–80.

[49] Kline JA, Hogg MM, Courtney DM, Miller CD, Jones AE, Smithline HA. D-dimer threshold increase with pretest probability unlikely for pulmonary embolism to decrease unnecessary computerized tomographic pulmonary angiography. *J Thromb Haemost*. 2012;10:572–81.

[50] Runyon MS, Beam DM, King MC, Lipford EH, Kline JA. Comparison of the Simplify D-dimer assay performed at the bedside with a laboratory-based quantitative D-dimer assay for the diagnosis of pulmonary embolism in a low prevalence emergency department population. *Emerg Med J*. 2008;25:70–5.

[51] Gibson NS, Sohne M, Gerdes VEA, Nijkeuter M, Buller HR. The importance of clinical probability assessment in interpreting a normal D-dimer in patients with suspected pulmonary embolism. *Chest*. 2008;134:789–93.

[52] Gibson NS, Douma RA, Squizzato A, Söhne M, Büller HR, Gerdes VEA. Application of a decision rule and a D-dimer assay in the diagnosis of pulmonary embolism. *Thromb Haemost*. 2010;103:849–54.

[53] Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant GJ. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med*. 2011;26:197–203.

[54] Barais M, Fossard E, Dany A, Montier T, Stolper E, Van Royen P. Accuracy of the general practitioner's sense of alarm when confronted with dyspnoea and/or chest pain: a prospective observational study. *BMJ Open*. 2020;10:e034348.

[55] Righini M, Le Gal G, Perrier A, Bounameaux H. The challenge of diagnosing pulmonary embolism in elderly patients: influence of age on commonly used diagnostic tests and strategies. *J Am Geriatr Soc*. 2005;53:1039–45.

[56] Engbers MJ, Vlieg AV, Rosendaal FR. Venous thrombosis in the elderly: incidence, risk factors and risk groups. *J Thromb Haemost*. 2010;8:2105–12.

SUPPLEMENTARY MATERIAL

The online version contains supplementary material available at https://doi.org/10.1016/j.jtha.2023.05.023