Extended D-dimer cut-offs and machine learning for ruling out pulmonary embolism in individuals undergoing computed tomography pulmonary angiography

Received: 22 Nov 2021
Accepted: 7 Feb 2022

To the Editor:

Pulmonary embolism (PE) is a major cause of morbidity and mortality [1]. Computed tomography pulmonary angiography (CTPA) is the gold standard for diagnosing PE [2] and a common investigation which contributes to potentially avoidable radiation exposure. CTPA use has quadrupled in the past two decades [3], and this has been associated with lower rates of PE detection [4] and possible overdiagnosis [5].

Despite efforts to make judicious use of CTPA, PE is typically only detected in 5–15% of scans [6]. Numerous clinical decision rules (CDRs) have been developed to aid clinicians, including the Wells [7] and Geneva scores [8], the Pulmonary Embolism Rule-out Criteria [9], YEARS [10] and PEGeD [11]. Clinician *gestalt* is heavily weighted in these CDRs. Conversely, D-dimer measurement is advised as a follow-up test, to be considered after CDRs have been applied, despite the fact that D-dimers below the upper limit of normal (ULN) are the most robust predictor of absence of PE, typically ruling out PE in ⩾98% of individuals [10, 11]. Furthermore, D-dimers increase with age, leading to the validation of "age-adjusted D-dimer" (aaD-dimer) thresholds [12, 13]. Recent studies suggest that extending D-dimer thresholds to 1 pg·mL^{-1} in low-risk individual, effectively rules out PE at 3-month follow-up [10, 11].

We hypothesised that combining D-dimers and risk factors in a model which removes the subjective likelihood of PE could provide robust PE rule-out performance. Furthermore, we sought to explore extended D-dimer thresholds to identify novel cut-offs for PE prediction in moderate-risk patients. We performed a single-centre, retrospective, proof-of-concept study to develop a PE rule-out algorithm. We trained a machine learning model for PE prediction in a PE-enriched training dataset (a discovery set of real-world consecutive scans, combined with a set of exclusively PE-positive scans to balance outcomes and improve model training), testing performance in a validation dataset of consecutive CTPAs.

Training and validation scans were performed between 1 January, 2017 and 30 May, 2020, and between 1 January and 31 December, 2016, respectively. Age and sex were documented. CTPA requests were reviewed to identify the presence of risk factors: Wells score components [7] (excluding "PE is the most likely diagnosis"), hormone replacement therapy/oral contraceptive pill, peri-partum status, chest pain, loss of consciousness and hypoxaemia. CTPA-reported PE was recorded as a binary outcome. CTPAs performed to confirm/monitor previously identified PEs and those deemed non-diagnostic were excluded. D-dimer level (HemosIL D-dimer HS 500 assay; Instrumentation Laboratory, Bedford, MA, USA) was documented if measured within 24 h prior to CTPA. aaD-dimer thresholds were calculated for patients >50 years old, using the formula (age/100) pg·mL^{-1}. Only data from CTPAs accompanied by D-dimers measured within the prior 24 h were included.

We assessed the performance of a gradient boosting classifier (*xgboost*), a generally high-performing algorithm for classification tasks, examining the role of D-dimer thresholds (ULN, 1.5×ULN, 2×ULN) in combination with Wells components and relevant co-variables as predictors in the model. A model training pipeline was created, with predictor variables assessed in several combinations using age, Wells score

Shareable abstract (@ERSpublications)
Combining novel machine learning algorithms with extended D-dimer cut-offs may improve pulmonary embolism prediction and reduce patient radiation exposure resulting from avoidable scans https://bit.ly/3oUKd7X

Cite this article as: Franciosi AN, McCarthy N, Gaffney B, *et al*. Extended D-dimer cut-offs and machine learning for ruling out pulmonary embolism in individuals undergoing computed tomography pulmonary angiography. *Eur Respir J* 2022; 59: 2200075 [DOI: 10.1183/13993003.00075-2022].

components and D-dimer. We proposed a model incorporating a given D-dimer threshold (*Θ*), where the decision rule of any model *M*, given a set of features *F*, and D-dimer threshold (*Θ*) would be:

$$f(M,\ F,\ \Theta) = \begin{cases} f(M,\ F)\ if\ \text{D-dimer} < \Theta \\ Predict\ PE\ otherwise \end{cases}$$

Performance and comparison to simple pre-defined D-dimer thresholds (ULN, 1.5×ULN, 2×ULN and aaD-dimer) was assessed in the validation cohort.

Statistical analysis was performed in R v4.0.4 (the R Foundation for Statistical Computing) and Jupyter Lab v2.2.6 running Python 3.8 (scikit-learn v0.24.1, pandas v1.2.3). All comparative tests were two-sided with p-values <0.05 considered significant.

Of 1047 CTPAs screened in the discovery dataset, 572 with D-dimers were included (mean age 44.6 years, 39.5% male). 2688 CTPAs were screened for inclusion in the PE enrichment group, with 367 were positive for PE (13.7%), of which 190 with D-dimers were included (mean age 47.3 years, 44.2% male). These two sets formed the PE enriched training set (n=762). 1314 scans were screened for the validation cohort and 634 (48.2%) with D-dimers were included (mean age 43.6 years, 36.8% male). PE prevalence did not differ between CTPAs included or excluded based on D-dimer availability (discovery cohort: 15.7% *versus* 14.3%, p=0.59; validation cohort: 14.5% *versus* 15.5%, p=0.69). PE prevalence did not differ between the discovery and validation cohorts (15.7% *versus* 15.5%, p=0.96). Median (interquartile range) D-dimer did not differ between discovery and validation cohorts (1.17 (0.74–2.24) pg·mL^{-1} *versus* 1.15 (0.70–2.40) pg·mL^{-1}; p=0.87) but was markedly higher in the enrichment (PE-positive) cohort (3.54 (1.76–7.10) pg·mL^{-1}; p <0.001).

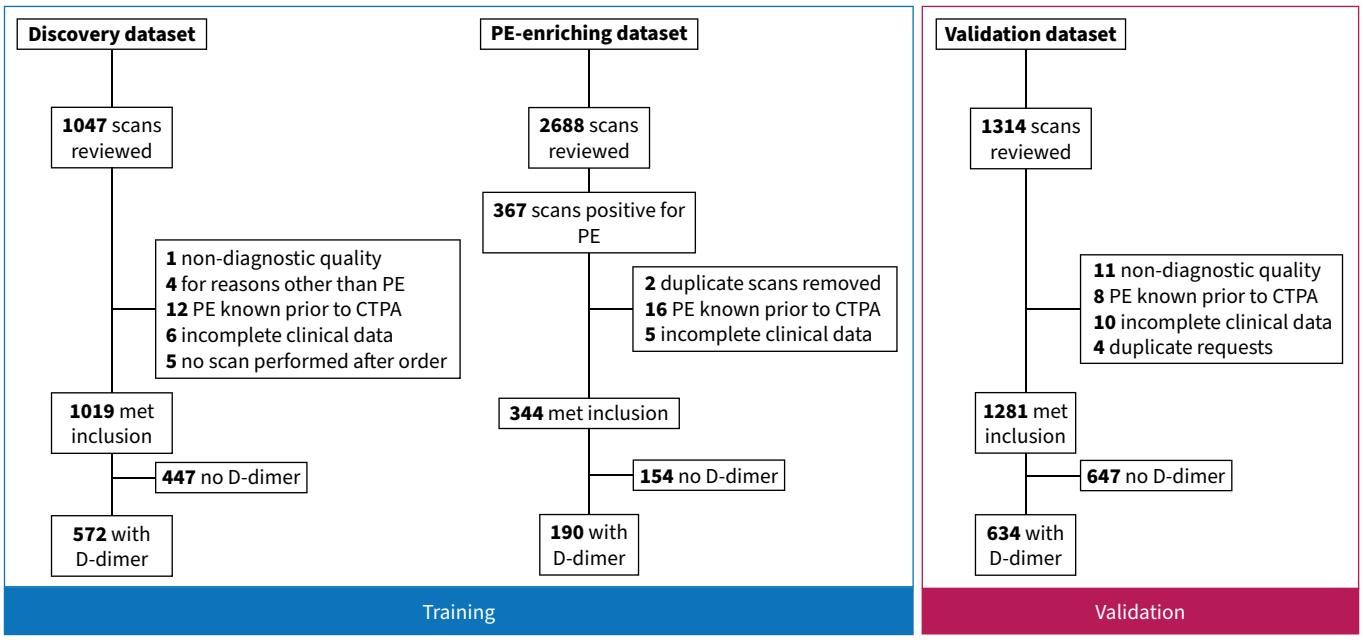
Among the models trained we found that a model incorporating a D-dimer threshold of 1.5×ULN (0.75 pg·mL^{-1}), Wells score components and age as predictors performed best in validation (negative predictive value (NPV) 99.3%, sensitivity 99.0%, specificity 27.4%). The performance of the model compared to the ULN and aaD-dimer for rule-out of PE was 99.3% *versus* 98.1% *versus* 98.2%, respectively for NPV (p=ns), 98.9% *versus* 98.9% *versus* 97.96%, respectively for sensitivity (p=ns), and 23% *versus* 8% *versus* 16%, respectively for scans predicted negative for PE (p<0.0001 by McNemar test) (figure 1).

We describe the results of a proof-of-concept study investigating novel approaches to PE prediction based on the analysis of 1396 CTPAs of individuals deemed clinically to be at least at moderate risk of PE who underwent CTPA scans during usual care. Using easily available clinical predictors we demonstrate that a gradient boost classifier (*xgboost*) model outperformed traditional and exploratory D-dimer thresholds for ruling out PE. This model achieved a high NPV (99.3%), sensitivity of 98.98%, and would outrule significantly more CTPAs than using ULN or aaD-dimer thresholds.

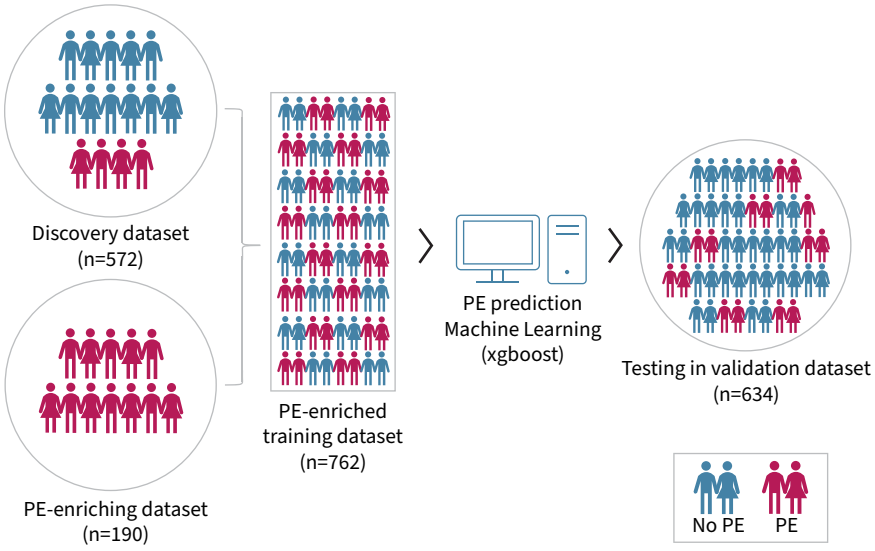
In 2017 the International Society on Thrombosis and Haemostasis (ISTH) Subcommittee on Predictive and Diagnostic Variables in Thrombotic Disease suggested that the historically accepted failure rate of 2.7% for venous thromboembolism may not be valid and that a lower failure threshold of 1.8–2% should be used to calculate power for future prospective studies [14]. Our model achieved the target NPV suggested by the ISTH and performed better than the ULN cut-off, tripling the number of CTPAs predicted negative.

Our study has some limitations, including the retrospective single-centre design, though the ∼15% PE prevalence in all cohorts suggests reasonable CTPA use. Secondly, CTPA outcome was determined by real-world reports, and not by re-interpretation of the images. Additionally, clinical factors were drawn from unstructured clinician CTPA requests, meaning the presence of a variable can be assumed to be reliable, but the absence of a risk factor in the request cannot as some features may have been inadvertently omitted by referring physicians. Moreover, only positive predictors of PE (*e.g.* malignancy, immobility), and not factors associated with a negative likelihood of PE, were considered. Collectively, these issues suggest that greater performance could be achieved in future prospective studies adopting similar approaches. Furthermore, while the included CTPAs were deemed clinically necessary during clinical care, systematic prospective risk scoring was not available and selection bias due to the exclusion of CTPAs without D-dimer must be considered. Nonetheless, the similarity in PE prevalence between scans included and the excluded cohorts suggests that no bias in PE risk was introduced by exclusion based on absence of D-dimer.

a)



b)



c)

		Performance (n=634)													
	D-dimer threshold	% predicted negative	% PEs missed	Sens	Spec	PPV	NPV	FNR	FPR	TNR	TPR	TP	FP	FN	TN
D-dimer alone	ULN	8.36	1.02	98.98	9.70	16.70	98.11	0.01	0.90	0.10	0.99	97	484	1	52
	Age-adjusted#	16.09	2.04	97.96	18.66	18.05	98.04	0.02	0.81	0.19	0.98	96	436	2	100
	1.5×ULN	28.55	4.08	95.92	33.02	20.75	97.79	0.04	0.67	0.33	0.96	94	359	4	177
	2×ULN	43.38	10.20	89.80	49.44	24.51	96.36	0.10	0.51	0.49	0.90	88	271	10	265
Machine learning¶	1.5×ULN	23.34	1.02	98.98	27.43	19.96	99.32	0.01	0.73	0.27	0.99	97	389	1	147
	2×ULN	35.96	7.14	92.86	41.23	22.41	96.93	0.07	0.59	0.41	0.93	91	315	7	221

FIGURE 1 a) Prisma diagram of study population selection. b) Study design: a discovery set of 572 consecutive computed tomography pulmonary angiograms (CTPAs) (pulmonary embolism (PE) prevalence 15.7%) was combined with a set of 190 exclusively PE-positive CTPAs to balance outcomes so as to improve classification training. This PE-enriched training set (n=762) was used to train models and performance was tested in

the validation dataset comprised of 634 consecutive CTPAs (PE prevalence 15.5%). c) Performance metrics of the simple D-dimer thresholds and machine learning models in the validation cohort. Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; FNR: false negative rate; FPR: false positive rate; TNR: true negative rate; TPR: true positive rate; TP: true positive; FP: false positive; FN: false negative; TN: true negative; ULN: upper limit of normal. #: age-adjusted D-dimer threshold, *i.e.* (age/100) pg·mL^{-1}; ¶: gradient boost classifier incorporating a given D-dimer cut-off, Wells score components and age.

Extending D-dimer cut-offs beyond the upper limit of normal may be applicable even in populations with moderate-to-high pre-test probability of PE, potentially extending the insights from YEARS and PeGed and improving pre-test-prediction and resource utilisation. These results provide insights into possible future of PE risk stratification strategies.

Alessandro N. Franciosi[1,2,7], Nicholas McCarthy[3,7], Brian Gaffney[1,4], John Duignan[4], Eamon Sweeney[1], Niall O'Connell[1], Karen Murphy[5], Fionnuala Ní Áinle[3,6], Marcus W. Butler[1,3], Jonathan D. Dodd[3,4], Michael P. Keane[1,3], David J. Murphy[3,4], Kathleen M. Curran[3] and Cormac McCarthy[1,3]

[1]Dept of Respiratory Medicine, St. Vincent's University Hospital, Dublin, Ireland. [2]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada. [3]School of Medicine, University College Dublin, Dublin, Ireland. [4]Dept of Radiology, St. Vincent's University Hospital, Dublin, Ireland. [5]Dept of Haematology, St Vincent's University Hospital, Dublin, Ireland. [6]Dept of Haematology, Mater Misericordiae Hospital, Dublin, Ireland. [7]Denotes joint first authorship.

Corresponding author: Cormac McCarthy (Cormac.McCarthy@UCD.ie)

Author contributions: A.N. Franciosi conceptualised the study and design, performed data analysis, wrote the manuscript and is the co-lead author. N. McCarthy co-designed the study, performed data analysis, performed machine learning analyses and is the co-lead author. B. Gaffney, J. Duignan, E. Sweeney and N. O'Connell performed data collection, preliminary data coding and participated in study design. K. Murphy, F. Ní Áinle, M.W. Butler, J.D. Dodd, M.P. Keane and D.J. Murphy consulted on study design, performed internal review and edited the manuscript. K.M. Curran and C. McCarthy participated in study design, performed internal review and edited the final manuscript. C. McCarthy is the senior author and had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of interest: The authors report no conflict of interests.

References

1 Rali PM, Criner GJ. Submassive pulmonary embolism. *Am J Respir Crit Care Med* 2018; 198: 588–598.
2 Perrier A, Roy PM, Sanchez O, *et al.* Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med* 2005; 352: 1760–1768.
3 Wang RC, Miglioretti DL, Marlow EC, *et al.* Trends in imaging for suspected pulmonary embolism across US health care systems, 2004 to 2016. *JAMA Netw Open* 2020; 3: e2026930.
4 Chong J, Lee TC, Attarian A, *et al.* Association of lower diagnostic yield with high users of CT pulmonary angiogram. *JAMA Intern Med* 2018; 178: 412–413.
5 Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Arch Intern Med* 2011; 171: 831–837.
6 Mountain D, Keijzers G, Chu K, *et al.* Correction: RESPECT-ED: rates of pulmonary emboli (PE) and sub-segmental PE with modern computed tomographic pulmonary angiograms in emergency departments: a multi-center observational study finds significant yield variation, uncorrelated with use or small PE rates. *PLoS One* 2017; 12: e0184219.
7 Wells PS, Anderson DR, Rodger M, *et al.* Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. *Thromb Haemost* 2000; 83: 416–420.
8 Le Gal G, Righini M, Roy PM, *et al.* Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med* 2006; 144: 165–171.
9 Kline JA. Diagnosis and exclusion of pulmonary embolism. *Thromb Res* 2018; 163: 207–220.
10 van der Hulle T, Cheung WY, Kooij S, *et al.* Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. *Lancet* 2017; 390: 289–297.

11 Kearon C, de Wit K, Parpia S, *et al.* Diagnosis of pulmonary embolism with D-dimer adjusted to clinical probability. *N Engl J Med* 2019; 381: 2125–2134.
12 Crop MJ, Siemes C, Berendes P, *et al.* Influence of C-reactive protein levels and age on the value of D-dimer in diagnosing pulmonary embolism. *Eur J Haematol* 2014; 92: 147–155.
13 Righini M, Van Es J, Den Exter PL, *et al.* Age-adjusted D-dimer cutoff levels to rule out pulmonary embolism: the ADJUST-PE study. *JAMA* 2014; 311: 1117–1124.
14 Dronkers CEA, van der Hulle T, Le Gal G, *et al.* Towards a tailored diagnostic standard for future diagnostic studies in pulmonary embolism: communication from the SSC of the ISTH. *J Thromb Haemost* 2017; 15: 1040–1043.