DIAGNOSTIC AND PROGNOSTIC RISK STRATIFICATION OF **VENOUS THROMBOEMBOLISM** IN PRIMARY CARE

Janneke Hendriksen

DIAGNOSTIC AND PROGNOSTIC RISK STRATIFICATION
OF VENOUS THROMBOEMBOLISM IN PRIMARY CARE

ISBN 978-94-6233-206-5
Cover and lay-out Teuntje van de Wouw
Printed by Gildeprint, Enschede, the Netherlands
Printed on FSC certified paper

The studies in this thesis were funded by the Netherlands Organisation for Scientific Research (ZonMw project no. 171002214). Financial support by the Julius Center for Health Sciences and Primary Care, the FNT (Federation of Dutch Thrombosis Services) and the SBOH for the publication of this thesis is gratefully acknowledged.

DIAGNOSTIC AND PROGNOSTIC RISK STRATIFICATION OF VENOUS THROMBOEMBOLISM IN PRIMARY CARE

DIAGNOSTISCHE EN PROGNOSTISCHE
RISICOSTRATIFICATIE VAN VENEUZE
TROMBO-EMBOLIEËN IN DE HUISARTSPRAKTIJK
(MET EEN SAMENVATTING IN HET NEDERLANDS)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 9 februari 2016 des middags te 2.30 uur

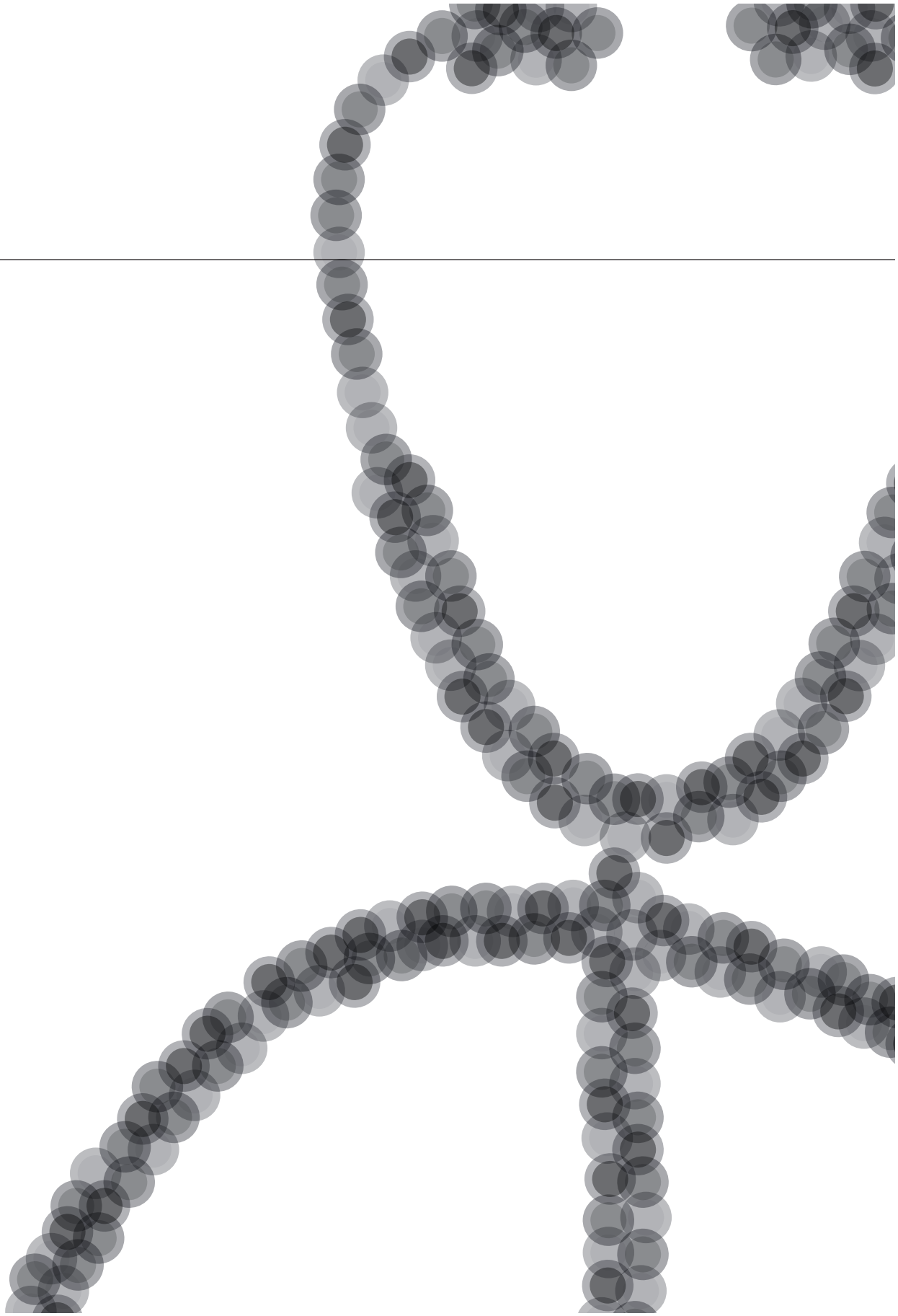
door

Johanna Maria Theresia Hendriksen
geboren op 9 juni 1986 te Nijmegen

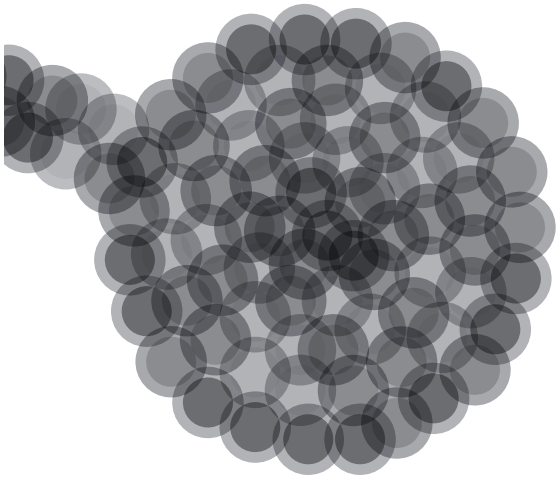
Promotor:	Prof. dr. K.G.M. Moons
Copromotoren:	Dr. G.J. Geersing
	Dr. R.E.G. Schutgens

Voor mijn ouders

CONTENTS



GENERAL INTRODUCTION



VENOUS THROMBOEMBOLISM

Venous thromboembolic disease is the result of an imbalance in the formation and degradation of fibrin in the venous blood circulation. Inappropriate fibrin formation leads to the development of thrombi or emboli, that can cause vein obstruction and concomitant complaints. The two main manifestation forms of venous thromboembolism (VTE) are deep vein thrombosis (DVT), usually in the lower extremities, and pulmonary embolism (PE), in which the emboli are located in the pulmonary arteries.

VTE is a highly clinically relevant disease in terms of the associated mortality risk and detrimental long-term health effects. The landmark study by Barritt and Jordan in the early 1960s reported a considerable mortality risk of 25% in patients with PE who were left untreated. (1) Currently, this mortality risk is still considered to be relevant as more recent observational cohorts and autopsy studies estimate it to be 10-30%, although others state that mortality is under 5% in ambulant patients. (2-4) In addition, it is estimated that yearly 500.000 individuals in Europe die because of the consequences of VTE, making it the third leading cause of cardiovascular death (after myocardial infarction and stroke). (5)

Long-term health effects after a VTE are also numerous. The risk of anticoagulant treatment associated bleeding increases with 0.5%/year in low risk patients, up to >4.0% per year in case of high risk. (6) Furthermore, there is an increased chance of recurrence and chronic complications like post thrombotic syndrome (PTS) (reported in 30-50% of patients post-DVT), and chronic thromboembolic pulmonary hypertension (CTEPH) in up to 3% in PE. (7-9) Both conditions are associated with substantial loss of quality of life and increased health care related costs. (10-13) Given these figures, the burden of VTE and its health impact are considerable and awareness of VTE, both by physicians and patients, is of much importance.

DIAGNOSTIC AND PROGNOSTIC RISK STRATIFICATION

General practitioners (GPs) play an important role in prompt diagnosis, management and counselling in venous thromboembolic disease. In the diagnostic phase, GPs have to judge the patient's condition and decide whether to refer for objective testing in secondary care, or alternatively refrain from referral if the risk of having VTE is considered very low. Once a diagnosis has been established, knowledge of the disease prognosis is important to be able to deal with questions on optimal treatment duration or counselling.

Both in disease diagnosis and prognosis, the key word is *risk estimation*: the probability of a VTE being present (diagnosis) or will (re)occur (prognosis) determines further actions. Inevitably, predictions are surrounded by a substantial amount of uncertainty and as a consequence, clinical dilemmas are frequently encountered in the diagnosis and prognosis of VTE.

CLINICAL DILEMMAS

DIAGNOSIS

The absolute number of patients that visit their general practitioner with signs and symptoms leading to the suspicion of VTE is relatively low: on average, each general practitioner comes across 14 patients with suspected DVT (10 cases) or suspected PE (4 cases) each year. (14)

The often non-specific presentation of this potentially life-threatening condition hampers the diagnostic process regularly. For example, coughing and mild dyspnea are frequently presented complaints at the general practitioner's office. In the majority of cases, these symptoms can be attributed to a respiratory tract infection. However, in a small subset of patients, coughing and mild dyspnea are caused by a pulmonary embolism instead. In fact, PE is present in only 1 out of 8 suspected cases and DVT in 1 out of 5 suspected cases. (15, 16) This frequently poses GPs with a diagnostic dilemma. Should they refer a patient with the least suspicion of DVT or PE ("just to be on the safe side"), therewith increasing the diagnostic burden in secondary care, while most referred patients will not have DVT or PE at all? Or will they refer only in case of a clear suspicion, but at the risk of missing a (less obvious) case of DVT or PE?

This decision is largely fueled by the estimated probability of the diagnosis being present. This probability can be estimated implicitly or explicitly. Gestalt ('gut feeling') is an important implicit diagnostic tool for physicians, especially in a primary care setting. In secondary care, multiple diagnostic prediction models have been developed to explicitly guide the risk stratification and are safe to use in combination with D-dimer testing. (17-20) It is currently still a matter of debate what strategy should be used by GPs to guide their referral decision, particularly if PE is suspected.

PROGNOSIS

The cornerstone of treatment of VTE consists of anticoagulant treatment. Its function is twofold: anticoagulants prevent thrombus growth and the development of recurrent events. VTE recurrence occurs frequently, especially in patients in whom the index event occurred in the absence of clear provoking factors such as recent major surgery. (21) Prevention is important, since VTE recurrence is associated with a higher risk of chronic complications like CTEPH or PTS. (7, 9) Although the effectiveness of anticoagulant therapy to prevent VTE is without doubt, the most important disadvantage of anticoagulant use is the increased risk of bleeding. The risk of major bleedings (that is: intracranial or gastrointestinal bleeding event) is approximately 1% per year, depending on risk factors like age and comorbidity. (22-24) Guidelines therefore advise to use anticoagulants for a defined period of 3-6 months for the majority of patients. (6, 25)

This one-size-fits-all approach might not be feasible in all patients. Ideally, the individual risk of VTE recurrence is an important factor to tailor one's treatment

duration, taking into account one's bleeding risk as well. To do so, it is important to identify the determinants that are associated with an increased VTE recurrence risk. Is it possible to predict one's individual risk of recurrence, and if so, can these predictions be used to tailor one's anticoagulant treatment duration on an individual level?

OBJECTIVES OF THESIS

Most knowledge on diagnostic and prognostic risk stratification in VTE is based on studies in secondary care. Although it is tempting to implement these diagnostic and prognostic strategies in a primary care setting as soon as possible, this should not be done without proper evaluation in this new setting first. Situation-specific factors are likely to influence the performance of these strategies, but the extent and direction of the influence can not always be predicted. Thus, testing of these strategies in a primary care domain is necessary and requires a structured approach.
In this thesis, the objectives are to evaluate the value of different risk stratification strategies to:

1. adequately identify VTE in primary care while restraining from over-referral (diagnosis of VTE)
2. tailor anticoagulant treatment duration in unprovoked VTE (prognosis of VTE)

OUTLINE OF THESIS

This thesis elaborates on the two aforementioned objectives. Difficulties in the *diagnosis* of VTE in primary care and potential tools to overcome them, are touched upon first. In **chapter 1** we illustrate the challenges GPs face in the diagnostic process of PE, based on a retrospective chart review. **Chapter 2** provides a comprehensive overview of developing diagnostic and prognostic models in the VTE domain. Next, the performance of five diagnostic models to rule out PE is evaluated for the primary care domain in **chapter 3**. In **chapter 4**, we compare the use of a structured diagnostic prediction model and the GP's 'gestalt' estimate in the diagnostic process of suspected PE. **Chapter 5** reports on the cost-effectiveness of using point-of-care D-dimer tests, on top of the use of a diagnostic prediction model, to rule out DVT in a primary care setting.

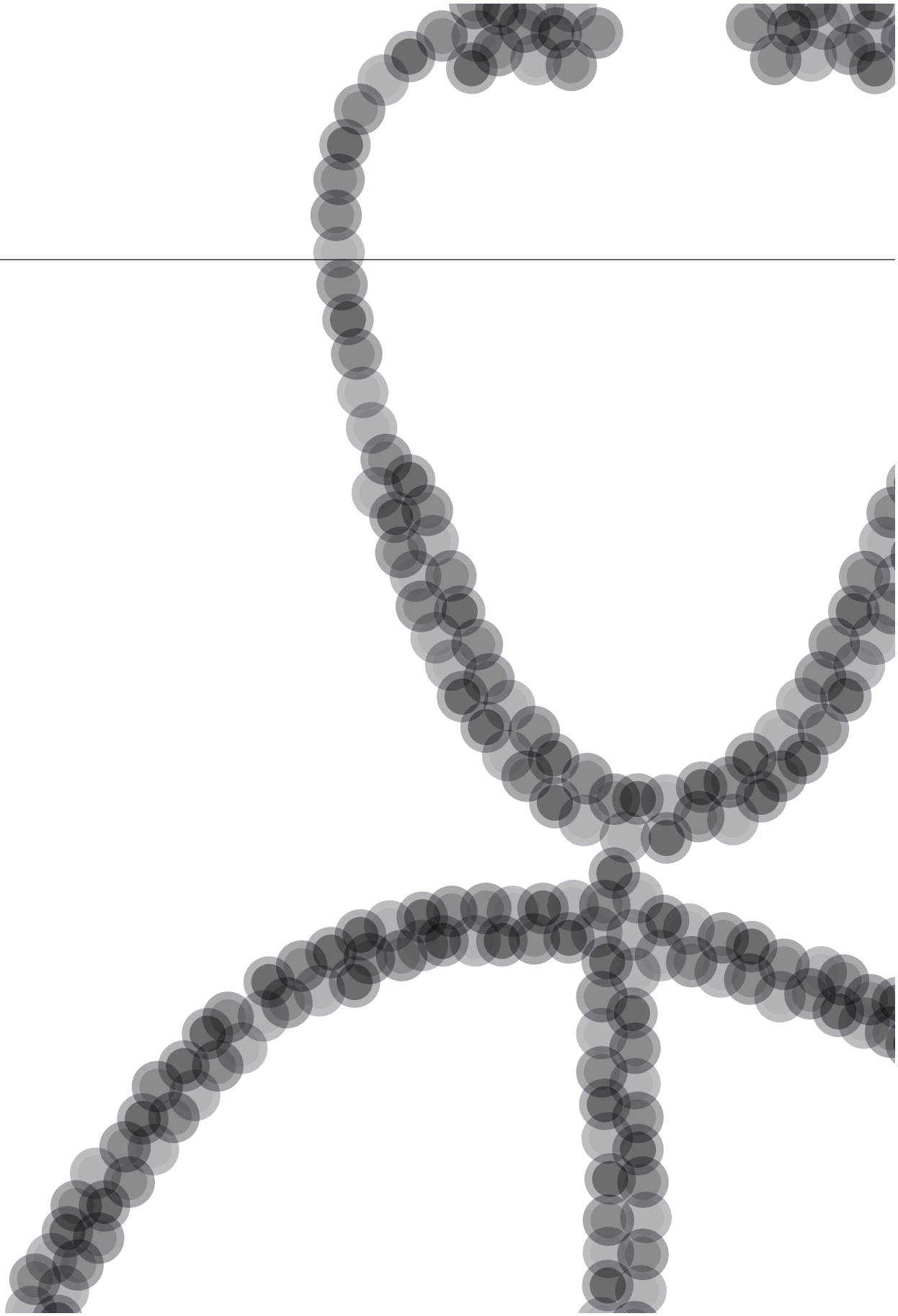
Once the diagnosis of venous thromboembolism has been established, the patient's and GP's focus shifts towards *prognosis* and subsequent treatment strategies. **Chapter 6** elaborates on the concept of tailored anticoagulant treatment duration based on the use of risk factors for recurrent venous thromboembolism, or combined in a prognostic prediction model, in patients with unprovoked VTE.

In **chapter 7** we present the design of the VISTA study, a pragmatic randomized trial on the use of a prediction model on recurrent VTE to tailor anticoagulant treatment duration in patients with a (first) unprovoked venous thromboembolic event. The main findings and the clinical implications of the thesis are summarized and commented upon in the general discussion.

REFERENCES

1. Barritt D, Jordan S. Anticoagulant drugs in the treatment of pulmonary embolism. A controlled trial. Lancet. 1960;275(7138):1309-12.
2. Kearon C. Natural history of venous thromboembolism. Circulation. 2003;107(23 suppl 1):I22-30.
3. Calder KK, Herbert M, Henderson SO. The mortality of untreated pulmonary embolism in emergency department patients. Ann Emerg Med. 2005;45(3):302-10.
4. Dalen J, Alpert J. Natural history of pulmonary embolism. Prog Cardiovasc Dis. 1975;17(4):259-70.
5. Cohen AT, Agnelli G, Anderson FA, Arcelus JI, Bergqvist D, Brecht JG, et al. Venous thromboembolism (VTE) in Europe - The number of VTE events and associated morbidity and mortality. Thromb Haemost. 2007;98(10):756-64.
6. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-based clinical practice guidelines. Chest. 2012;141(2_suppl):e419S-e94S.
7. Kahn SR, Shrier I, Julian JA, Ducruet T, Arsenault L, Miron MJ, et al. Determinants and time course of the postthrombotic syndrome after acute deep venous thrombosis. Ann Intern Med. 2008;149(10):698-707.
8. Hoeper MM, Madani MM, Nakanishi N, Meyer B, Cebotari S, Rubin LJ. Chronic thromboembolic pulmonary hypertension. Lancet Respir Med. 2014;2(7):573-82.
9. Pengo V, Lensing AWA, Prins MH, Marchiori A, Davidson BL, Tiozzo F, et al. Incidence of chronic thromboembolic pulmonary hypertension after pulmonary embolism. N Engl J Med. 2004;350(22):2257-64.
10. Klok FA, Zondag W, van Kralingen KW, van Dijk APJ, Tamsma JT, Heyning FH, et al. Patient outcomes after acute pulmonary embolism. Am J Respir Crit Care Med. 2010;181(5):501-6.
11. van Korlaar I, Vossen C, Rosendaal F, Cameron L, Bovill E, Kaptein A. Quality of life in venous disease. Thromb Haemost. 2003;90(7):27-35.
12. van Es J, den Exter PL, Kaptein AA, Andela CD, Erkens PMG, Klok FA, et al. Quality of life after pulmonary embolism as assessed with SF-36 and PEmb-QoL. Thromb Res. 2013;132(5):500-5.
13. Lefebvre P, Laliberté F, Nutescu EA, Duh MS, LaMori J, Bookhart BK, et al. All-cause and disease-related health care costs associated with recurrent venous thromboembolism. Thromb Haemost. 2013;110(12):1288-97.
14. Baaten G, van Bakel H, Geersing G, de Jong J, Lucassen W, Van Noortwijk H, et al. Diepe veneuze trombose en longembolie (eerste herziening). Huisarts Wet. 2015;58(1):26-35.
15. Geersing GJ, Erkens PMG, Lucassen W, Buller H, Ten Cate H, Hoes A, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
16. Oudega R, Moons KG, Hoes AW. Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. Thromb Haemost. 2005;94(1):200-5.

17. Wells P, Andersen D, Rodger M, Ginsberg JS, Kearon C, Gent M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. Thromb Haemost. 2000;83(3):416-20.
18. Le Gal G, Righini M, Roy P-M, Sanchez O, Aujesky D, Bounameaux H, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. Ann Intern Med. 2006;144(3):165-71.
19. 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(1):92-7.
20. Kline JA, Mitchell AM, Kabrhel C, Richman PB, Courtney DM. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. J Thromb Haemost. 2004;2(8):1247-55.
21. Prandoni P, Noventa F, Ghirarduzzi A, Pengo V, Bernardi E, Pesavento R, et al. The risk of recurrent venous thromboembolism after discontinuing anticoagulation in patients with acute proximal deep vein thrombosis or pulmonary embolism. A prospective cohort study in 1,626 patients. Haematologica. 2007;92(2):199-205.
22. Palareti G, Leali N, Coccheri S, Poggi M, Manotti C, D'Angelo A, et al. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). Lancet. 1996;348(9025):423-8.
23. Torn M, Bollen W, van der Meer F, van der Wall E, Rosendaal F. Risks of oral anticoagulant therapy with increasing age. Arch Intern Med. 2005;165(13):1527-32.
24. Ruiz-Gimenez N, Suarez C, Gonzalez R, Nieto JA, Todoli JA, Samperiz AL, et al. Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry. Thromb Haemost. 2008;100(1):26-31.
25. Konstantinides SV, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galie N, et al. 2014 ESC Guidelines on the diagnosis and management of acute pulmonary embolism. Eur Heart J. 2014;35:3033-80.



CHAPTER 1

CLINICAL CHARACTERISTICS ASSOCIATED WITH DIAGNOSTIC DELAY OF PULMONARY EMBOLISM IN PRIMARY CARE

Janneke MT Hendriksen, Marleen L Koster- van Ree,
Marcus J Morgenstern, Ruud Oudega, Roger EG Schutgens,
Karel GM Moons, Geert-Jan Geersing

Submitted.

1

ABSTRACT

Rationale: Pulmonary embolism (PE) is one of the most frequently missed diagnoses in clinical practice as symptoms can be relatively mild or non-specific, mimicking other common cardiopulmonary diseases. Given the unselected patient population and the limited availability of diagnostics on the spot, delay or missing of a diagnosis is expected to be present in primary care medicine as well. Delay of PE in this specific domain however is hardly studied. Hence, the objective of this study was to evaluate the extent of delay in PE diagnosis in primary care, and the determinants associated with such diagnostic delay.

Methods: We performed a retrospective observational study in six primary care practices. All PE cases (ICPC-code K93) confirmed by imaging up to June 2015 were extracted from the electronic medical records. Next, for all these PE events we reviewed all consultations with their general practitioner (GP) and scored any signs and symptoms that could be attributed to PE in the 3 months prior to the event. Also, we documented actual comorbidity and the diagnosis considered initially. Delay was defined as a time gap of >7 days between the first contact with the GP and the final diagnosis. Multivariable logistic regression analysis was performed to identify independent determinants for delay.

Results: In total 180 incident PE cases were identified, of whom 128 patients had one or more potential PE-related contact with their GP within the three months prior to the diagnosis. Based on our definition, in 33 of these patients (26%) diagnostic delay was observed. Older age (age >75 years) (OR 5.10 (95%CI 1.84-14.13)) and the absence of chest complaints (that is: chest pain or pain on inspiration) (OR 5.37 (95%CI 1.90-15.16)) were independent determinants for diagnostic delay. A respiratory tract infection prior to the PE diagnosis was reported in 14% of all cases, and in 33% of patients with delay.

Conclusion: Diagnostic delay of more than seven days in the diagnosis of pulmonary embolism is common in primary care, especially in elderly and if chest complains, like pain on inspiration, are absent.

INTRODUCTION

Pulmonary embolism (PE) is listed among the diagnoses most frequently missed, or delayed, in clinical practice. (1) In fact, a substantial part of the estimated 500,000 deaths per year attributed to venous thromboembolism in Europe are likely to be contributed to a missed or delayed diagnosis. (2) This delay is thought to be driven primarily by a non-specific disease presentation. The classical triad of dyspnoea, pain on inspiration and haemoptysis is only present in 10% of the patients with PE. (3) Instead, PE presentation can range from complaints mimicking a simple cough or myalgia, an acute myocardial infarction or even nephrolithiasis. This non-specific presentation poses a major diagnostic challenge to physicians to identify PE timely. Prompt recognition of the diagnosis, and consequent immediate initiation of anticoagulant therapy, is important to prevent severe complications and mortality, and is recommended by current guidelines. (4, 5)

The evidence on appropriate strategies to diagnose PE is overwhelming, yet all these strategies by definition start with a suspicion of PE. In contrast, limited evidence is available on the magnitude of delayed PE diagnoses and determinants associated with such diagnostic delay. The current evidence comes largely from studies, performed in emergency departments (EDs). These studies identified that diagnostic delay was common, yet with varying proportion of delayed PE cases (range 12% to 75% of PE cases). (6-13) Furthermore, the definitions used to quantify delay were heterogeneous. Various determinants like higher age, comorbidity (e.g. chronic obstructive pulmonary disease (COPD), cardiovascular disease), absence of dyspnoea and no pain on respiration were associated with a delay in diagnosis, but these findings were not consistent across studies.

In many countries general practitioners (GPs) fulfil a pivotal gate-keeping role for access to subsequent hospital care. Yet, GPs have only limited diagnostic tools available on the spot and often have to rely substantially on clinical assessment. Appreciating the relatively high reported number of PE events missed or delayed in a pre-selected ED population and the diagnostic limitations in primary care, we hypothesize that delay is likely to be present frequently in primary care as well. However, the extent of delay in PE diagnosis in primary care, and its related determinants, are currently unknown. Knowledge on these determinants might help GPs to better identify patients with PE in time.

Therefore, the two aims of the present study were to explore the extent of diagnostic delay in primary care, and to identify determinants that are associated with this diagnostic delay.

METHODS

STUDY DESIGN

We performed an observational study in six primary care practices in the

Netherlands. The study was assessed by the local Institutional Ethics Review Board and received a waiver for formal reviewing. As such, according to Dutch law, no explicit informed consent was required as data reducible to the patients was only available at the GPs practices and made anonymous for data evaluation and analysis by the researchers.

STUDY POPULATION

All patient contacts labelled with the International Classification of Primary Care (ICPC)-code K93 (i.e. diagnosis of pulmonary embolism) until June 2015 were extracted from the electronic patient records (EPR) of the six practices. Next, detailed information on all consultations three months prior to the PE diagnosis were scrutinized using the following approach. First, all patients with an ICPC-code K93 were validated on the actual presence of PE by the researchers (MM, MvR) using hospital discharge information, including results from imaging and medication prescriptions. In case of doubt another reviewer was consulted (JH, GJG). Patients were excluded from further analysis if PE was considered absent, if insufficient information on the PE event was available to confirm the diagnosis, if no objective imaging was performed (e.g. in palliative patients) or if the ICPC-code was assigned incorrectly in the EPR (e.g. if PE was suspected initially, but ruled out after referral and objective imaging). Additionally, the diagnosis PE was confirmed if anticoagulant therapy was initiated post-event and if the hospital discharge letter described the presence of PE based on diagnostic imaging tests.

OUTCOME: DIAGNOSTIC DELAY

We assessed all GP contacts prior to the final diagnosis PE on their relevance in relation to the final PE diagnosis. Patient contacts were considered to be relevant for this PE event when the patient had presented with any of the following signs or symptoms: dyspnoea, cough, haemoptysis, chest pain, painful respiration, fever (body temperature >38°C), increased respiratory rate, increased heart rate, low oxygen saturation, or signs of deep venous thrombosis (DVT). If data on these items were not reported, we considered them absent. In all other cases (e.g. regular diabetes work-up, psychosocial problems) we tagged these contacts as not diagnostically relevant to the final PE diagnosis. Additionally, the initial diagnosis suspected by the GP at the time of the first patient contact, relevant for the final PE event, was registered in all cases. Also the number of days between the first presentation of any of our pre-defined signs or symptoms relevant for PE at the GP and the final PE diagnosis was calculated. Cases in which there was any doubt regarding relevance of GP contacts were discussed during consensus meetings (MM, MvR, JH and GJG).

The outcome, delay in PE diagnosis, was defined as a period longer than 7 days between the patient's first GP contact attributed to be relevant for PE (based on our aforementioned definition) and the final PE diagnosis, similar to a previous study on diagnostic delay. (10) Immediate referral to the emergency department based on the suspicion of another condition than PE was not considered to be

delay, since the severity of the condition was acknowledged.

POTENTIAL DETERMINANTS OF DIAGNOSTIC DELAY

The following determinants were assessed to be a potential determinant for diagnostic delay: older age (>75 years), gender, risk factors for PE (i.e. recent immobilization, prior venous thromboembolism, systemic oestrogen use, pregnancy, puerperium and recent surgery), medical history (i.e. COPD, asthma, hypertension, coronary artery disease, heart failure, atrial fibrillation, (past) smoking and malignancy (based on ICPC coding and/ or referral letters as recorded in the EPR)). Finally, results of D-dimer testing were assessed.

STATISTICAL ANALYSES

The determinants were first compared between patients with and without a delay in diagnosis in a univariable analysis. Continuous variables were presented as mean (standard deviation (SD)) or median (interquartile range (IQR)) and compared with the independent sample T-test or the Mann-Whitney U test. Categorical variables were reported as the absolute number (plus percentage) and compared using the Chi-square test or Fisher's Exact test.

We then performed multivariable logistic regression analysis using presence or absence of diagnostic delay as the binary outcome to assess which of the potential determinants were independently associated with delayed PE diagnosis. First, we constructed a logistic model using signs and symptoms, age and gender as potential determinants for diagnostic delay. Next, this logistic model was extended with co-morbidity in a second model, to gain further insight into the type of patients in whom a delay of diagnosis occurs more often in primary care medicine. Regression coefficients from the logistic models were recalculated into odds ratios with their surrounding 95% confidence interval.

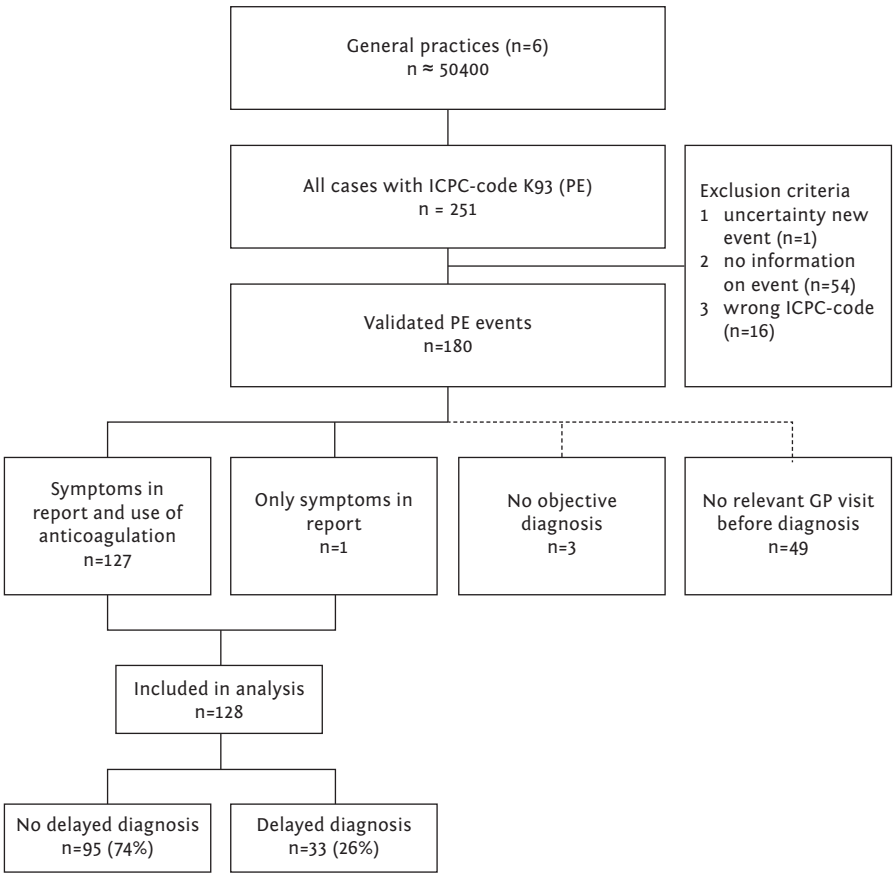
Given the exploratory nature of this study as well as the uncertainty around the scope of diagnostic delay for PE in primary care we deliberately chose not to define statistical criteria regarding sample size. Instead, acknowledging the fact that our study sample would not allow for a selection of variables into the logistic models based on p-values, we a priori defined the variables that we wanted to assess in the logistic models. This variable selection was entirely based upon previous literature review, and included gender, age, absence of chest complaints (i.e. pain on inspiration and chest pain), and the absence of dyspnoea for the first logistic model, and additionally a prior respiratory chest infection and asthma/COPD for the second logistic model. Data were analysed using SPSS 21.0 (SPSS Inc, Chicago, IL).

RESULTS

Approximately 50,400 patients were registered in the six general practices that participated. Using the ICPC-code K93, 251 possible pulmonary embolism cases were identified until June 2015. Seventy-one cases were excluded based on the predefined

1

FIGURE 1 FLOW CHART OF THE SELECTION OF PULMONARY EMBOLISM CASES IN PRIMARY CARE



PE= pulmonary embolism; ICPC= international classification of primary care; GP= general practitioner

exclusion criteria. For the majority of cases (n=54) no detailed information was available, for example if the event took place years ago while registered in another practice. In total, 180 verified PE cases were left for further analyses (see *Figure 1*). Forty-nine patients (27%) had no contact with their GP prior to the PE event and in three (palliative care) patients no objective imaging was performed. See *Table 1* for further characteristics of the group of patients that did not contact their GP in the period prior to the PE diagnosis. This latter group included more patients with a recent surgery, active malignant disease, but less patients with cardiac or respiratory comorbidities. The remaining 128 patients had relevant contact moments with their GP prior to the diagnosis and were included in our main analysis. In 33 of these

TABLE 1 CHARACTERISTICS OF PATIENTS WITH AND WITHOUT GP VISIT BEFORE DIAGNOSIS PE

	No prior GP visit (n=49)	Prior GP contact(s) (n=128)	p-value
Age in years***, mean (± SD)	60 (17)	58 (16)	0.429
Male gender	27 (55)	60 (47)	0.327
Comorbidity			
COPD/ asthma	5 (10)	27 (21)	0.067*
Hypertension	11 (22)	37 (29)	0.387
Atrial fibrillation	1 (2)	4 (3)	1.000*
Congestive heart failure	0 (0)	6 (5)	0.189*
Ischemic heart disease	3 (6)	14 (11)	0.406*
Risk factors			
History of DVT	7 (14)	26 (20)	0.357
Recent surgery	8 (16)	8 (6)	0.045*
Recent immobilization	4 (8)	6 (5)	0.467*
Recent travelling	2 (4)	4 (3)	0.669*
Malignancy	9 (18)	10 (8)	0.042
Estrogen use, women	0/22 (0)	19/68 (28)	0.005*
Prior pulmonary infection	3 (6)	23 (18)	0.057*
History of smoking	6 (12)	33 (26)	0.052
Pregnancy	0/22 (0)	3/68 (4)	1.000
Days until diagnosis, median [IQR]	0 [0]	1 [8]	
Total range	NA	0-126 days	
Delay > 7 days	NA	33 (26)	

Values are numbers (percentages) unless stated otherwise. *Fisher's exact test *** missing values (n=3)
PE= pulmonary embolism; GP= general practitioner; SD= standard deviation; COPD= chronic obstructive pulmonary disease; DVT= deep venous thrombosis; IQR= interquartile range; NA= not applicable.

patients (26%), diagnostic delay as defined a-priori was observed.

Patients with a delay in diagnosis were on average older, had less frequent chest pain (24% vs. 54%, p = 0.003) and less frequent pain on respiration (9% vs. 33%, p= 0.011) on initial presentation. Diagnosis was more often delayed in patients with a recent respiratory tract infection (33% vs. 13%, p= 0.008) (see *Table 2*). In the first multivariable logistic regression analysis, older age (OR 5.10 (95%CI 1.84-14.13)) and the absence of chest complaints (OR 5.37 (95%CI 1.90-15.16)) were associated with diagnostic delay (see *Table 3a*). In the second model, female gender, absence of dyspnoea and prior respiratory tract infections were associated with delay in diagnosis too (see *Table 3b*).

TABLE 2 BASELINE CHARACTERISTICS *(DIAGNOSTIC DELAY >7 DAYS)*

	No diagnostic delay (n=95)	**Diagnostic delay > 7 days** (n=33)	**p-value**
Age in years***, mean (± SD)	56 (15)	62 (18)	0.068
Male gender	47 (49)	13 (39)	0.317
Symptoms			
Dyspnea	62 (65)	22 (67)	0.884
Chest complaints	59 (62)	9 (27)	0.001
Chest pain	51 (54)	8 (24)	0.003
Painful respiration	31 (33)	3 (9)	0.011*
Cough	17 (18)	12 (36)	0.029
Hemoptysis	3 (3)	1 (3)	1.000*
Signs of DVT	18 (19)	1 (3)	0.025*
Fever (>38°C)	5 (5)	2 (6)	1.000*
Collapse	4 (4)	0 (0)	0.572*
Heart rate, mean (± SD)	96 (19)	95 (26)	0.916
O2 saturation, median [IQR]	96 [5]	94 [9]	0.124
Comorbidities			
COPD and/or asthma	14 (15)	13 (39)	0.004
Hypertension	25 (26)	12 (36)	0.273
Atrial fibrillation	2 (2)	2 (6)	0.273*
Congestive heart failure	3 (3)	3 (9)	0.177*
Ischemic heart disease	10 (11)	4 (12)	0.755*
Risk factors			
History of DVT	18 (19)	8 (24)	0.515
Recent surgery	7 (7)	1 (3)	0.679*
Recent immobilization	6 (6)	0 (0)	0.338*
Recent travelling	4 (4)	0 (0)	0.572*
Malignancy	7 (7)	3 (9)	0.717*
Estrogen use, women	15/48 (31)	4/20 (20)	0.393*
Prior pulmonary infection	12 (13)	11 (33)	0.008
History of smoking	26 (27)	7 (21)	0.486
Pregnancy	3/48 (6)	0/20 (0)	0.550

Values are numbers (percentages) unless stated otherwise. *Fisher's exact test *** missing values (n=3) PE= pulmonary embolism; SD= standard deviation; DVT= deep venous thrombosis; IQR= interquartile range; COPD= chronic obstructive pulmonary disease; NA= not applicable.

TABLE 3A MULTIVARIABLE LOGISTIC REGRESSION MODEL FOR THE ASSOCIATION BETWEEN SIGNS AND SYMPTOMS WITH A DIAGNOSTIC DELAY > 7 DAYS

MODEL 1	OR (95%CI)	p-value
Female gender	2.47 (0.93 to 6.61)	0.071
Age >75	5.10 (1.84 to 14.13)	0.002
No chest complaints	5.37 (1.90 to 15.16)	0.002
No dyspnea	2.29 (0.81 to 6.46)	0.118

OR= odds ratio; 95%CI= 95% confidence interval.

TABLE 3B MULTIVARIABLE LOGISTIC REGRESSION MODEL FOR THE ASSOCIATION BETWEEN SIGNS, SYMPTOMS AND COMORBIDITY WITH A DIAGNOSTIC DELAY >7 DAYS

MODEL 2	OR (95%CI)	p-value
Female gender	3.08 (1.08 to 8.78)	0.036
Age >75	4.28 (1.49 to 12.28)	0.007
No chest complaints	5.35 (1.79 to 16.05)	0.003
No dyspnea	3.08 (1.01 to 9.35)	0.047
Prior respiratory tract infection	3.34 (1.11 to 10.01)	0.031
COPD/ asthma	3.34 (0.88 to 7.54)	0.085

OR= odds ratio; 95%Ci= 95% confidence interval.

DISCUSSION

In this study on the extent of diagnostic delay of pulmonary embolism in primary care and its determinants, we observed that diagnostic delay was present in a substantial proportion of patients (26%). An important factor associated with diagnostic delay appeared to be the absence of the typical "text book" chest pain symptoms as presented during the first presentation at the general practitioner. Furthermore, in those with diagnostic delay, a respiratory tract infection was frequently reported, before the initial PE diagnosis was established.

This is the first study on the magnitude of diagnostic delay of pulmonary embolism in a primary care setting. Strengths of this study are the fact that we had access to the general practitioner's electronic patients records, including all consultations with the GP prior to the diagnosis, plus the correspondence between GPs and hospital specialists and all results from laboratory and imaging tests performed. This allowed us to sketch a complete and detailed picture of the diagnostic pathway starting from the first presentation of patients with signs and symptoms at their GP to the final diagnosis of PE.

However, for full appreciation of these results, some limitations need to be addressed. Foremost, we used a retrospective design to quantify diagnostic delay, which incurs several challenges.

First, we had to rely on correct ICPC-coding in all primary care practices. Only cases with the ICPC-code K93 were extracted for detailed assessment, leaving the chance of erroneously leaving out PE cases that were labelled with an incorrect ICPC-code. This could be the case if a patient is referred to secondary care with non-specific complaints like dyspnoea, consequently coded as such, without updating of the coding to PE afterwards. We however believe that this miscoding will be present only in a fraction of all PE cases, given that PE is an important diagnosis to be reported for further anticoagulant use and because of its prognostic implications. Another reason for incomplete selection of PE cases is the fact that patient records of those deceased were not available for all. This leaves the chance of missing PE events in patients that occurred in the years prior to their passing.

Second, data on determinants like oxygen saturation, heart rate and D-dimer levels were not reported in the patient records of many cases. For the current analyses, we assumed that the variable was not present if information on that item was not reported. However, it cannot be said for sure if these variables were not measured at all, or absent if not reported. Especially if PE is not suspected by the GP, it can be expected that not all PE risk factors and signs (e.g. recent traveling or respiratory rate) are explicitly asked. If so, this can lead to selective underreporting of specific determinants. In an attempt to gain insight into the possible selective reporting, we performed a sensitivity analysis in which we tested for the distribution of non-reported factors between the patients with and without delay. No differences were observed, however (data not shown).

Third, we had to make interpretations on the relevance of certain complaints for the diagnosis PE, all with current knowledge that PE was indeed present. Since blinding is not possible in a retrospective design, no measures could be taken to prevent this.

Nevertheless, being aware of all these drawbacks of using a retrospective study design, we deemed this design to be the most suitable method to study delay in diagnosis. By definition, delay is only to be determined with hindsight and as a consequence, prospective evaluation of the diagnostic process will be difficult for a disease with a relatively low incidence in primary care.

We arbitrarily defined diagnostic delay as a time lag of >7 days between the first presentation at the GP and the final diagnosis. However, a lag period of >7 seven days can be deemed to be too large, especially considering the potential fatal outcome of not treating a PE event in time. (5) Therefore, we performed a sensitivity analysis in which we set the definition of diagnostic delay at 1 or more GP contact without adequate referral to secondary care for further diagnostics. Using this (more stringent) definition of diagnostic delay, the proportion of cases missed increased to over 40% (data not shown). Yet, the inferences on our reported determinants of

delay remained constant, strengthening the validity of our findings.

Another complicating factor in the definition of delay in PE diagnosis concerns its physiological development and progression. In recent years, much has been written on the interplay of inflammation and an activated thrombogenic state. (14, 15) Thus, an infection could also be the precursor of, and provoking factor for, PE. An initial diagnosis of a respiratory tract infection that turns out to be a PE a few days later is not necessarily a delayed PE diagnosis. Instead, the PE can rather be the result of a cascade initiated by the infection. This distinction between a PE initiated by an infection, and a PE incorrectly diagnosed as an infection, cannot be made easily, leaving the chance of overestimation of the association between a respiratory tract infection and diagnostic delay. This possible interplay between infection and venous thrombosis was also observed the other way around in a recent study by Timp and colleagues. (16) Patients receiving antibiotics (as a proxy for infectious disease) had an increased risk of both first and recurrent VTE (incidence rate ratio (IRR) 5.6 (95% CI 4.6-6.8)). Analogue to our inferences, this point estimate may have been too high due to misclassification of PE symptoms as an infection, and thus inappropriate prescription of antibiotics. The latter in fact would be classified as delay in diagnosis in our study.

In seven PE cases without diagnostic delay, the diagnosis of PE was not considered initially by the GP. However, due to the suspicion of another serious condition (acute coronary event), urgent referral to the emergency department led to prompt diagnosis of PE anyhow. In the strict sense of the definition, the diagnosis PE is missed here since it was not the diagnosis deemed to be most likely by the GP. We treated these missed, but adequately referred, cases as "delay present" in a sensitivity analysis, not changing our inferences (data not shown).

Finally, we did not collect data on the clinical implications of diagnostic delay, like quality of life or long-term health effects. However, knowledge on the impact of delay on clinical outcomes is important to be able to value the relevance of delayed diagnoses: it can be hypothesized that long-term outcomes are worse for patients who have had a prolonged duration of complaints, caused by the lengthened exposure to vessel obstruction, and consequent vessel damage, pulmonary hypertension or lung infarction. Further prospective research however is needed to address this research question for the primary care domain.

COMPARISON WITH OTHER STUDIES

A few studies on delay in diagnosis of pulmonary embolism have been performed in secondary care, especially in emergency departments. For example, Torres-Macho et al. found that patients with diagnostic delay in an emergency department were older, had more frequently COPD or asthma, and presented more often with a cough, fever and tachycardia. (8) In the study by Alonso- Martinez and colleagues, delay was also more prevalent amongst elderly, and in those with previous cardiac disease and without sudden onset of dyspnoea. (6) In our study, older age was associated with diagnostic delay too, just like the absence of typical PE complaints.

However, female gender was not reported previously in relation with delay. Only one study mentioned female gender as a risk factor for patient-delay rather than doctor-delay. (13) Given our small sample size and exploratory nature, these findings require further evaluation in future research.

CLINICAL IMPLICATIONS

Given the substantial percentage of cases with delay in diagnosis of PE observed in this current study, we assume that most GPs come across a situation with diagnostic delay in PE. Even if an alternative diagnosis is confirmed (e.g. infiltrate on chest radiography, or an initial improvement after the initiation of antibiotic treatment), PE can be present simultaneously, or develop along the course of the concurrent disease. Therefore, we argue that an initial diagnosis should be reconsidered frequently, especially if symptoms do not improve as much as can be expected, and in those with coexisting cardiopulmonary conditions.

Furthermore, we showed in this current study that the diagnosis of PE is more often delayed in case of non-typical presentation of complaints. Therefore, we suggest GPs to consider PE as a potential diagnosis even if symptoms are not overwhelmingly pointing into that direction. To prevent an overshoot of referrals to secondary care as a consequence of this low threshold of suspicion, a diagnostic prediction model might help GPs to further guide the decision whether or not to refer a patient. (17) Nevertheless, especially in elderly and those with concurrent respiratory tract infections, the false-positive rate of these prediction models (specifically the D-dimer test) is considerable. (18, 19) As such, the optimal balance between raised awareness of PE and refraining from over-referral has yet to be found and further evaluation of diagnostic delay is needed.

CONCLUSION

Diagnostic delay of pulmonary embolism is common in primary care, especially if classical PE symptoms like chest complaints are absent. Awareness of the possibility of a PE being the underlying cause of a wide range of symptoms, might contribute to a reduction of the number of delayed PE diagnoses in primary care.

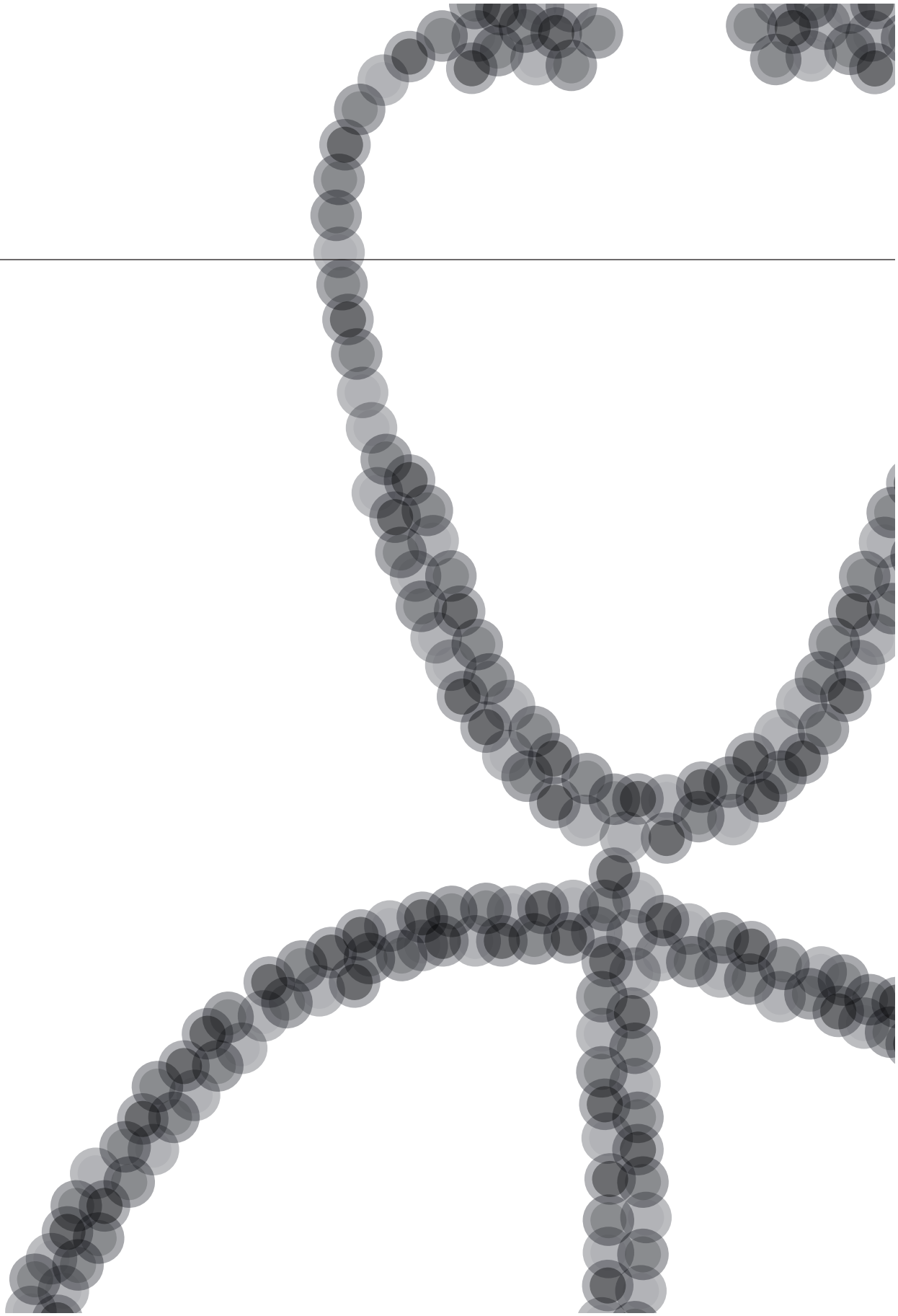
ACKNOWLEDGEMENTS

We would like to thank the following general practices and GPs for their participation in this project: Praktijk Buitenhof, Amsterdam; S. van Doorn MD, Praktijk Spechtenkamp, Maarssen; F. Rutten MD PhD, Huisartsenpraktijk de Grebbe, Rhenen; J. Morgenstern MD, Huisartsenpraktijk Heerde; R. Damoiseaux MD PhD, Hof van Blom, Hattem; C. Hoppenreijs MD, Gezondheidscentrum Binnenstad, Utrecht.

REFERENCES

1. Schiff G, Hasan O, Kim S, Abrams R, Cosby K, Lambert B, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169(20):1881-7.
2. Cohen AT, Agnelli G, Anderson FA, Arcelus JI, Bergqvist D, Brecht JG, et al. Venous thromboembolism (VTE) in Europe - The number of VTE events and associated morbidity and mortality. Thromb Haemost. 2007;98(10):756-64.
3. Meyer G, Roy P-M, Gilberg S, Perrier A. Pulmonary embolism. BMJ. 2010;340.
4. NICE. Diagnosis and management of venous thromboembolic diseases. NICE quality standard 29. Manchester: National Institute for Health and Clinical Excellence; 2013.
5. Douketis JD, Kearon C, Bates S, Duku EK, Ginsberg JS. Risk of fatal pulmonary embolism in patients with treated venous thromboembolism. JAMA. 1998;279(6):458-62.
6. Alonso-Martínez JL, Sánchez FJA, Echezarreta MAU. Delay and misdiagnosis in sub-massive and non-massive acute pulmonary embolism. Eur J Intern Med. 2010;21(4):278-82.
7. Smith SB, Geske JB, Morgenthaler TI. Risk factors associated with delayed diagnosis of acute pulmonary embolism. J Emerg Med. 2012;42(1):1-6.
8. Torres-Macho J, Mancebo-Plaza AB, Crespo-Giménez A, de Barros MRS, Bibiano-Guillén C, Fallos-Martí R, et al. Clinical features of patients inappropriately undiagnosed of pulmonary embolism. Am J Emerg Med. 2013;31(12):1646-50.
9. Kline JA, Hernandez-Nino J, Jones AE, Rose GA, Norton HJ, Camargo CA. Prospective study of the clinical features and outcomes of emergency department patients with delayed diagnosis of pulmonary embolism. Acad Emerg Med. 2007;14(7):592-8.
10. den Exter PL, van Es J, Erkens PMG, van Roosmalen MJG, van den Hoven P, Hovens MMC, et al. Impact of delay in clinical presentation on the diagnostic management and prognosis of patients with suspected pulmonary embolism. Am J Respir Crit Care Med. 2013;187(12):1369-73.
11. Ozsu S, Oztuna F, Bulbul Y, Topbas M, Ozlu T, Kosucu P, et al. The role of risk factors in delayed diagnosis of pulmonary embolism. Am J Emerg Med. 2011;29(1):26-32.
12. Ageno W, Agnelli G, Imberti D, Moia M, Palareti G, Pistelli R, et al. Factors associated with the timing of diagnosis of venous thromboembolism: Results from the MASTER registry. Thromb Res. 2008;121(6):751-6.
13. Aydogdu M, Dogan N, Sinanogly N, Oguzulgen I, Demircan A, Bildik F, et al. Delay in diagnosis of pulmonary thromboembolism in emergency department: is it still a problem? Clin Appl Thromb Hemost. 2013;19(4):402-9.
14. Delvaeye M, Conway EM. Coagulation and innate immune responses: can we view them separately? Blood. 2009;114(12):2367-74.
15. Tichelaar YIGV, Kluin-Nelemans HJC, Meijer K. Infections and inflammatory diseases as risk factors for venous thrombosis. A systematic review. Thromb Haemost. 2012;107(5):827-37.
16. Timp J, Cannegieter S, Tichelaar V, BRaekkan S, Rosendaal F, le Cessie S, et al. Risk of first and recurrent venous thrombosis in individuals treated with antibiotics. XXV Congress of the International Society on Thrombosis and Haemostasis; June 20–25, 2015; Toronto2015. p. 983.

17. Geersing GJ, Erkens PMG, Lucassen W, Buller H, Ten Cate H, Hoes A, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
18. Schouten HJ, Geersing GJ, Koek HL, Zuithoff NPA, Janssen KJM, Douma RA, et al. Diagnostic accuracy of conventional or age adjusted D-dimer cut-off values in older patients with suspected venous thromboembolism: systematic review and meta-analysis. BMJ. 2013;346.
19. Harper PL, Theakston E, Ahmed J, Ockelford P. D-dimer concentration increases with age reducing the clinical value of the D-dimer assay in the elderly. Intern Med J. 2007;37(9):607-13.

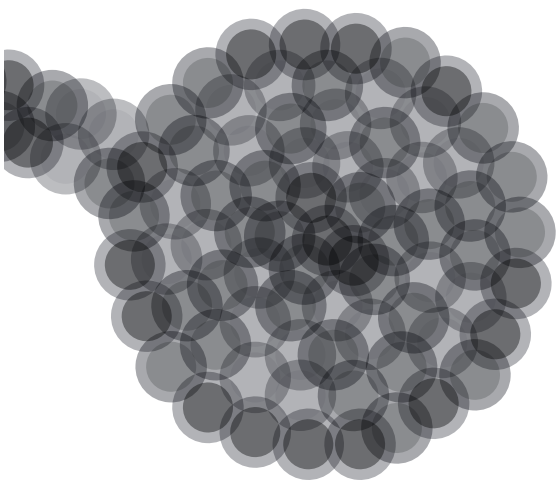


CHAPTER 2

DIAGNOSTIC AND PROGNOSTIC PREDICTION MODELS IN VENOUS THROMBOEMBOLISM

Janneke MT Hendriksen, Geert-Jan Geersing,

Karel GM Moons, Joris AH de Groot



J Thromb Haemost 2013; 11 (Suppl. 1): 129–41.

SUMMARY

Risk prediction models can be used to estimate the probability of either having (diagnostic model) or developing a disease (prognostic model). In clinical practice, these models are used to guide patient management. Examples from the field of venous thromboembolism (VTE) include the Wells rule for patients suspected of deep venous thrombosis and pulmonary embolism, and more recently prediction rules to estimate the risk of recurrence after a first episode of unprovoked VTE. In this paper, the three phases that are needed before a prediction model can be used in daily practice are described: development, validation, and implementation. In the development phase, the focus is on model development using either a logistic (diagnostic) or cox (prognostic) multivariate regression analysis. Model performance is checked by the discrimination and calibration of the final model. Discrimination refers to the ability to distinguish diseased from non-diseased, whereas calibration refers to the agreement of the predictive probability with the actual frequency of the outcome. In the validation phase, these same model performance measures are again checked in a new set of patients. This is important, as model performance is almost always poorer in a new set of patients, e.g. due to domain differences (primary or secondary care). Finally, in the implementation phase the ability of prediction model to actually guide patient management is evaluated. Whereas in the development and validation phase observational designs are preferred, this last phase often asks for randomized designs where patient management is compared using either the prediction model or usual care.

INTRODUCTION

In recent years, risk prediction models have become increasingly popular to aid clinical decision making. These models are developed to provide for estimating a probability of having (a diagnostic prediction model) or developing (a prognostic prediction model) a certain outcome (e.g. disease, event, complication) in an individual, given the individual's demographics, test results, or disease characteristics. The probability estimates can guide care providers as well as the individuals themselves in deciding upon further management. (1-4) In the field of venous thromboembolism (VTE), well known prediction models are those developed by Wells and colleagues. These rules aid the diagnostic process in patients suspected of deep venous thrombosis (DVT) or pulmonary embolism (PE) (see *Table 1*). (5, 6) Yet, many more prediction models in the domain of VTE have been developed, such as the prognostic models to assess VTE recurrence risk in patient who suffered from a VTE (7-9) or the Pulmonary Embolism Severity Index (PESI) for short-term mortality risk in PE patients (10), and various other diagnostic models for both DVT and PE, for example developed by Oudega et al (11) or Aujesky et al (10).

With the increase in risk prediction models developed and reported each year, the methodology for developing, validating and implementing these models receive increasingly attention, as reflected in recent books and series of publications. (4, 12-22) Unfortunately, the quality of a prediction model is not guaranteed by its publication as reflected by various recent reviews. (23-27) The very recent PROGRESS series reviews common shortcomings in model development and reporting. (22) The fact that multiple prediction models are being developed for a single clinical question, outcome or target population, suggests that there is still a tendency towards developing more and more models, rather than to first validate those existing or adjust an existing model to new circumstances. As a consequence, there is still a huge mismatch between the number of papers on model development versus on validation and even more versus the implementation of prediction models. (22, 28-30)

In this paper, we review the literature on methods for developing, validating and assessing the impact of prediction models, building on three recent series of such papers. (4, 14-18, 31) We illustrate this throughout with examples from the diagnostic and prognostic VTE domain, complemented with empirical data on a diagnostic model for PE. We stress that the empirical data, based on a recent publication of a model validation study of the Wells PE rule (6) for suspected PE in primary care (32), are used for illustration purposes only, and by no means to define the best diagnostic model or work-up for PE suspicion or to compare our results with existing reports on the topic. Our sole aim here is to illustrate the methods used in prediction modeling to improve understanding and interpretation of such studies.

TABLE 1 RISK PREDICTION MODELS (CLINICAL DECISION RULES) IN DIAGNOSIS OF DVT AND PULMONARY EMBOLISM

	Items	Regression coefficients	Points assigned
Wells DVT	Active cancer	-	1
(5)	Immobilization	-	1
	Recent surgery	-	1
	Tenderness	-	1
	Entire leg swollen	-	1
	Unilateral calf swelling >3 cm	-	1
	Unilateral edema	-	1
	Collateral superficial veins	-	1
	Alternative diagnosis as likely or more likely	-	-2

DVT unlikely (score ≤1 and low D-dimer): 33% (1268/ 3875) (90)
DVT present if DVT unlikely: 0.7% (95% CI 0.3%-1.3%) (90)

	Items	Regression coefficients	Points assigned
Wells PE	Clinical signs & symptoms of DVT	1.8	3.0
(6)	Other diagnosis less likely	1.5	3.0
	Heart rate > 100 bpm	1.1	1.5
	Recent immobilization or surgery	0.92	1.5
	Previous DVT or PE	0.87	1.5
	Hemoptysis	0.87	1.0
	Active cancer	0.81	1.0

PE unlikely (score ≤4 and low D-dimer): 42% (95% CI 33%-52%)(39)
PE present if PE unlikely: 1.7% (95% CI 1.0%-2.8%)(39)

RISK PREDICTION MODELS: DEFINITION AND RATIONALE

Risk prediction models estimate the risk (absolute probability) of the presence or absence of an outcome or disease in individuals based on their clinical and non-clinical characteristics. (1-3, 12, 33, 34) Depending on the amount of time until outcome assessment, prediction research can be diagnostic (outcome or disease present at this moment) or prognostic (outcome occurs within a specified time frame). Although we illustrate some of our methods with empirical data of a diagnostic modeling study, the methods described in this paper for prediction model development, validation and impact assessment, can be mutatis mutandis applied to both situations. (18)

In clinical diagnostic practice, doctors incorporate information from history taking, clinical examination, laboratory or imaging test results, to judge and determine whether or not a suspected patient has the targeted disease. In essence, prediction model development mimics this diagnostic work-up by combining all this patient information, further summarized as predictors of the outcome, in a statistical multivariable model. (2, 12, 33, 35-38)

For each unique combination of predictors, a prediction model provides an estimated probability that allows for risk stratification for individuals or groups. Hence, it can guide physicians in deciding upon further diagnostic tests or treatments. For example, patients with a high probability of having a disease might be suitable candidates for further testing, while in low probability patients it might be more effective to refrain from further testing. For instance, the combination of the Wells PE rule and a negative D-dimer test can safely rule out PE in about 40% of all patients suspected of having PE. These patients can be refrained from further testing, thus improving efficiency of the diagnostic process. (39)

The urge to develop a prediction model usually starts with a clinical question on how to tailor further management considering the patients profile of having or developing a certain outcome or disease. For example, patients with unprovoked VTE might benefit from prolonged anticoagulant therapy, but only those at high-risk for recurrence because of the associated risk of bleeding. Several tools have been developed to assess the prognostic probability of developing a recurrent VTE to determine whether or not secondary prevention is indicated in a subset of patients. (7-9, 40)

As addressed previously, to become clinically valuable, a prediction model ideally follows three clearly distinct steps, namely: development, validation and impact/ implementation. (12, 14, 18, 22, 28, 34)

DEVELOPMENT

DESIGN OF DATA COLLECTION

Ideally, the data needed to develop a new prediction model come from a prospective study, performed in study participants that share most of the clinical characteristics with the target patients for the model (i.e. generalizability of the model). (3) In diagnostic model development this means that a sample of patients suspected of having the disease are included, whereas the prognostic model requires subjects that might develop a specific health outcome over a certain time period. For example, the prognostic VTE recurrence prediction models were developed from prospective cohorts of VTE patients being at risk of a recurrent event. (7-9, 40)

Randomized clinical trials (RCTs) are in fact more stringently selected prospective cohorts. Data from RCTs can thus also be used for prognostic model development, yet – given the stringent inclusion and exclusion criteria – there is a chance of hampered generalizability. (14, 18) In contrast, a retrospective cohort design is prone to incomplete data collection as information on the predictors and outcomes is commonly less systematically obtained and therefore more prone to yield biased prediction models. The traditional case-control design is hardly suitable for risk prediction model development (and validation). However, a nested case-control or case-cohort design can be chosen for prediction modeling studies in specific circumstances, like a rare outcome or expensive predictor measurements. (41-43)

PREDICTORS

Many patient-related variables (i.e. sex, age, comorbidities, severity of disease, test results) that are known or assumed to be related with the targeted outcome may be studied as a predictor. Out of all such potential predictors, a selection of the most relevant candidate predictors has to be chosen to be included in the analyses. Especially when the number of subjects with the outcome is relatively small, as we will describe below (see *Table 2* and *3*: of all characteristics of patients suspected of DVT, we chose to include only seven predictors in our analyses). In contrast to etiologic study designs, in which only causally related variables are considered, non-causal variables can also be highly predictive of outcomes. (14) For example, one of the predictors of the Wells diagnostic PE rule is tachycardia (see *Table 2* and *3*). Although there is no causal relation between tachycardia and PE, the predictive ability is substantial.

Predictors that are difficult to measure, or have high inter-observer variability, might not be suitable for inclusion in a prediction model because this will influence the predictive ability of the model when applied in other individuals. A subjective predictor like "other diagnosis less likely" of the Wells PE rule might be scored differently by residents and more experienced physicians. Furthermore, it is of utmost importance to define predictors accurately and to describe the

measurements in a standardized way. This enhances applicability and predictive stability across multiple populations or settings of the prediction model to be developed. (33)

Continuous predictors (such as the D-dimer level in the Vienna Prediction Model (8), blood pressure or weight) can be used in prediction models, but preferably should not be presented as a categorical variable. Converting the variable into categories often creates a huge loss of information. (44, 45) Moreover, chosen thresholds for categorization are usually driven by the development data at hand, making the developed prediction model unstable and less generalizable when used or applied in other individuals. Continuous predictors should thus be kept continuous although it is important to assess the linearity or shape of the predictor- outcome association, and to transform the predictor if necessary. (13, 16, 44-46)

The decision on what candidate predictors to select for the study aimed at developing a prediction model is mainly based on prior knowledge, clinical or from the literature. Preferably, predictor selection should not be based on statistical significance of the predictor-outcome association in the univariable analysis (12, 13, 47, 48) (see also section on actual modeling). Also, it is often tempting to include as many predictors as possible into the model development. But if the number of outcome events in the dataset is limited, there is a high chance of including predictors into the model erroneously, only based on chance. (12, 13, 47, 48) To prevent this, although not based on firm scientific evidence, one might apply as a rule of thumb the so-called "EPV (events per variable) 1 to 10": one candidate predictor per 10 outcome events should be included in the dataset to secure reliable prediction modeling. (49-51) Other methods to limit the amount of candidate predictors are to combine several related variables into one single predictor or to remove candidate predictors that are highly correlated with others. (13)

OUTCOME

The outcome of a prediction model has to be chosen as such that it reflects a clinically significant and patient relevant health state, for example death yes or no, or absence or presence of (recurrent) pulmonary embolism. A clear and comprehensive pre-defined outcome definition limits the potential of bias. This includes a proper protocol on standardized (blinded or independent) outcome assessment. (4) In case of prognostic prediction research, a clear-defined follow-up period is needed in which the outcome development is assessed. For example, the PESI score, developed to identify PE patients with a low risk of short-term mortality in whom outpatient treatment may be safe, used 30 days of follow-up to assess the outcome PE recurrence or mortality during that period. (10)

MISSING DATA

As in all types of research, missing data on predictors or outcomes are unavoidable

TABLE 2 UNIVARIABLE ANALYSES OF EACH CANDIDATE DIAGNOSTIC PREDICTOR COMPARED TO THE PRESENCE OR ABSENCE OF PULMONARY EMBOLISM

	Pulmonary embolism (reference standard)					
	yes n=73			no n=525		
	N	**Sens**	**PPV**	**N**	**Spec**	**NPV**
Clinical signs and symptoms DVT	26	36 (26-47)	46 (33-58)	31	94 (92-96)	91 (89-93)
PE most likely diagnosis	61	84 (73-90)	18 (15-23)	272	48 (44-52)	95 (92-97)
Heart rate > 100 beats/ min	25	34 (24-46)	23 (16-31)	86	84 (80-87)	90 (87-92)
Recent immobilization or surgery	23	32 (22-43)	24 (17-34)	71	86 (83-89)	90 (87-92)
Previous DVT or PE	18	25 (16-36)	21 (14-31)	66	87 (84-90)	89 (86-92)
Hemoptysis	5	7 (3-15)	24 (11-45)	16	97 (95-98)	88 (85-91)
Presence of malignancy	5	7 (3-15)	19 (9-38)	21	96 (94-97)	88 (85-91)
Positive result D-dimer test	70	96 (89-99)	19 (9-38)	219	58 (54-62)	99 (97-100)

DVT = deep venous thrombosis; PE = pulmonary embolism; 95% confidence intervals between brackets.

To illustrate the development steps of a risk prediction model, we use data from a study in which the Wells PE rule was validated in a primary care setting. 598 patients suspected of having pulmonary embolism were included in the analysis. History, clinical examination and a dichotomous D-dimer test were performed in all participants. See reference (32) for a detailed discussion on all study logistics. Presence or absence of the outcome pulmonary embolism (PE) was assessed by a composite reference standard, including spiral CT scanning, V/Q scanning, angiography and 3 months follow-up. Although heart rate and d-dimer concentration are continuous predictors, they were analyzed here as dichotomous predictors conform the original definition in the diagnostic model (heart rate) and the dichotomous test used in the study (D-dimer). It can be seen that all single predictors are not sufficient to diagnose or reject the diagnosis pulmonary embolism. Of note, these univariable analyses are not used to select predictors to include in the multivariable model.

Calculation of the diagnostic accuracy measures is as follows:
Sens = sensitivity. The proportion of all diseased patients correctly classified as such based on the predictor.
Spec = specificity. The proportion of all non-diseased patients that is correctly classified as such based on the predictor.
PPV= positive predictive value. The probability of patients having PE given the fact that the predictor classified the patient as such.
NPV= negative predictive value. The probability of patients not having PE given the fact that the predictor classified the patient as not having PE.

TABLE 3 MULTIVARIABLE DIAGNOSTIC MODEL TO CONFIRM OR REJECT THE DIAGNOSIS PULMONARY EMBOLISM

	Pulmonary Embolism					
	Model 1 (Basic model)			Model 2 (Basic model + D-dimer)		
	regression coefficient (SE)	OR (95% CI)	p-value	regression coefficient (SE)	OR (95% CI)	p-value
(intercept)	-3.75 (0.34)	-	-	-5.88 (0.66)	-	-
Clinical signs and symptoms DVT	1.93 (0.33)	6.9 (3.6-13.2)	<0.01	1.50 (0.36)	4.5 (2.2-9.0)	<0.01
PE most likely diagnosis	1.32 (0.34)	3.8 (1.9-7.3)	<0.01	1.23 (0.36)	3.4 (1.7-6.9)	<0.01
Heart rate > 100 beats/ min	0.90 (0.31)	2.4 (1.3-4.5)	<0.01	0.56 (0.33)	1.7 (0.9-3.3)	0.09
Recent immobilization or surgery	0.71 (0.32)	2.0 (1.1-3.8)	0.03	0.61 (0.35)	1.8 (0.9-3.6)	0.08
Previous DVT or PE	0.91 (0.34)	2.5 (1.3-4.8)	<0.01	0.88 (0.37)	2.4 (1.2-5.0)	0.02
Positive result D-dimer test	NA	NA	NA	3.11 (0.61)	22.3 (6.8-73.1)	<0.01

DVT = deep venous thrombosis; PE = pulmonary embolism; OR = odds ratio; CI = confidence interval; SE = standard error; NA = not applicable.

The model development started with 7 candidate predictors of the Wells PE rule (see Table 2), plus D-dimer for quantifying its added value (see Table 4). According to the EPV rule 1 out of 10, the number of predictors should not be higher to prevent an overfitted model, as the total number of PE cases was 73 (see Table 2). Only 5 were included in the final model 1 when using a backward selection procedure using the AIC: in this dataset, active malignancy and hemoptysis did not have any independent value that was not already captured by the other predictors. The intercept reflects the baseline risk. The regression coefficient reflects the relative weight per predictor. The exponent of a regression coefficient yields the odds ratio (OR) of the predictor.

An OR of 2.0 for the predictor "Recent immobilization or surgery" in Model 1 indicates that the chance of having PE in a patient suspected of PE with is twice as high if the predictor is present, compared to a situation in which the predictor is absent, all other predictors kept constant.

To illustrate the effects of adding a new diagnostic biomarker to an existing prediction model (Model 1), we present a second model. Addition of a dichotomous D-dimer test (Model 2) yields regression coefficients that differ from the first model. Changed regression coefficients reflect that the history taking and physical examination predictors are correlated with the D-dimer results, but that the D-dimer test contributes most to the predicted probability.

For each individual, a predicted probability of having PE can be calculated using the formula:
Probability of PE = exp (lp)/ (1+exp(lp)).
"lp" stands for 'linear predictor' and is calculated by adding the baseline risk and the sum of all predictors multiplied by its regression coefficient. For example, for a patient where the physician considers the diagnosis of PE to be most likely, and a heart rate of 120 beats/ minute, while all other variables are absent (no signs and symptoms of DVT, no recent surgery, no previous DVT or PE), the prediction model 1 (without D-dimer) yields the following result:
lp= -3.75+ (0x 1.93) + (1x 1.32) + (1x 0.90) + (0x0.71) + (0x0.91) = -5.54
pPE = exp(-5.54)/(1+ exp (- 5.54)) = 17.8%.
The probability of having PE for this particular patient is 18%.

in prediction research as well. (52, 53) This can influence the model development, as missing data frequently follow a selective pattern in which the missingness of predictor results is related to other variables or even the outcome. (54-59) Removal of all participants with missing values is not sensible, since the non-random pattern of missing data inevitably causes a non-desired non-random selection of the participants with complete data as well. Moreover, it reduces the effective sample size. As a consequence, the model will be prone to inaccurate- biased- and attenuated effect size estimations.

Guidelines recommend imputing these missing data using imputation techniques. (55, 58-61) These techniques use all available information of a patient—and that of similar patients—to estimate the most likely value of the missing test results or outcomes in patients with missing data.

A predictor with many missing values, however, suggests difficulties in acquiring data on that predictor, even in a research setting. In clinical practice that specific variable will likely be frequently missing as well and one might argue if it is prudent to add such a predictor in a prediction model.

ACTUAL MODELING

Prediction models are usually derived using multivariable regression techniques, and many books and papers have been written how to develop a prediction model. (12, 13, 16, 62) In brief, a binary outcome commonly asks for the use of a logistic regression model for diagnostic or short-term (e.g. 1 or 3 months) prognostic outcomes, or survival modeling for long-term, time-to-event prognostic outcomes. There are two generally accepted strategies to arrive at the final model, yet there is no consensus on the optimal method to use. (12-14, 16)

The full model approach includes all candidate predictors not only in the multivariable analysis but also in the final prediction model, i.e. no predictor selection whatsoever is applied. The main advantage of this approach is a bypass of improper predictor selection due to chance (predictor selection bias). (13) The difficulty remains however to adequately preselect the predictors for inclusion in the modeling, and requires much prior knowledge. (16, 17)

To overcome this issue, the second method uses predictor selection in the multivariable analyses, either by backward elimination of 'redundant' predictors or forward selection of 'promising' ones. The backward procedure (see *Table 3*) starts with the full multivariable model (all predictors included, accounting for the above addressed "EPV 1:10 rule") and then subsequently removes predictors based on a predefined criterion, e.g. the Akaike Information Criterion (AIC) or a nominal significance level (based on the so-called Log likelihood ratio test (LR test)), for removal. (12) If predictors are added to the multivariable model one by one, this is called forward selection. With this approach, however, some variables will not be considered at all and thus no overall effect (that is, the model with all candidate predictors) is assessed. If selection is applied, backward selection is therefore often preferred over forward selection. Strict selection (e.g. based on an often used

significance level of <0.05) will lead to a low number of predictors in the final model but also enhances unintentional exclusion of relevant predictors, or inclusion of spurious predictors that by chance were significant in the development data set. This risk increases when the data set was relatively small and/or the number of candidate predictors relatively large. (12, 13, 18) Conversely, using less stringent exclusion criteria (e.g. p <0.25) leaves more predictors, but potentially also less important ones, in the model.

The predictors of the final model, regardless the selection procedure used, are considered all associated with the targeted outcome, yet the individual contribution to the probability estimation varies. The multivariable modeling assigns the weight of each predictor, mutually adjusted for each other's influence, to the probability estimate. For each individual the probability of having or developing the outcome can then be calculated based on these regression coefficients (see legend *Table 3*).

Developed regression models – logistic, survival or other - might be too complicated for (bedside) use in daily clinical care. To improve user-friendliness, the coefficients are often rounded towards numbers that can be easily scored by clinicians (see *Table 1*, Wells PE score). Such simplification, however, might hamper the accuracy of the model and thus needs to be applied with care. (18) Instead, one may use the original regression equation to create an easy to use web-based tool or nomogram to calculate individual probabilities. As an example of such comprehensive model presentation in the VTE domain, we refer to the Vienna Prediction Model nomogram and web-based tool. (8)

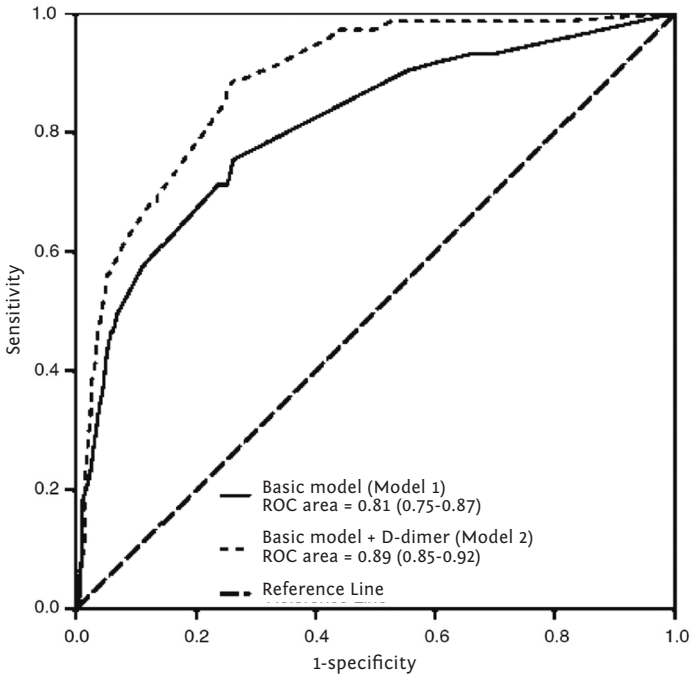
MODEL PERFORMANCE MEASURES

A prediction model should be able to distinguish diseased from non-diseased individuals correctly (discrimination) and should produce predicted probabilities that are in line with the actual outcome frequencies or probabilities (calibration). As the clinical potential of a prediction model largely depends on these two items, both should be assessed and reported as part of the model development.

Discrimination can be expressed as the area under the receiver-operating curve for a logistic model, or the equivalent c-index in a survival model. The AUC (or c-index) represents the chance that in two individuals, one with and one without the outcome, the predicted outcome probability will be higher for the individual with the outcome compared to the one without (see *Figure 1*). A c-index of 0.5 represents no discriminative ability, whereas 1.0 indicates perfect discrimination. (33, 63, 64)

A calibration plot provides insight into this calibrating potential of a model. First, individuals are ranked based on increasing model-derived deciles of predicted probability of having the outcome. On the x-axis the mean predicted probability and on the y-axis the observed outcome frequencies are plotted (see *Figure 2*). If the slope of a line equals 1 (diagonal), it reflects optimal calibration. A formal statistical test examines the so-called 'goodness-of-fit'. The Hosmer and Lemeshow test is regularly used, but might lack statistical power to detect overfitting. (12, 13, 65)

FIGURE 1 RECEIVER OPERATING CHARACTERISTIC CURVES FOR THE MODEL WITHOUT AND WITH D-DIMER TESTING



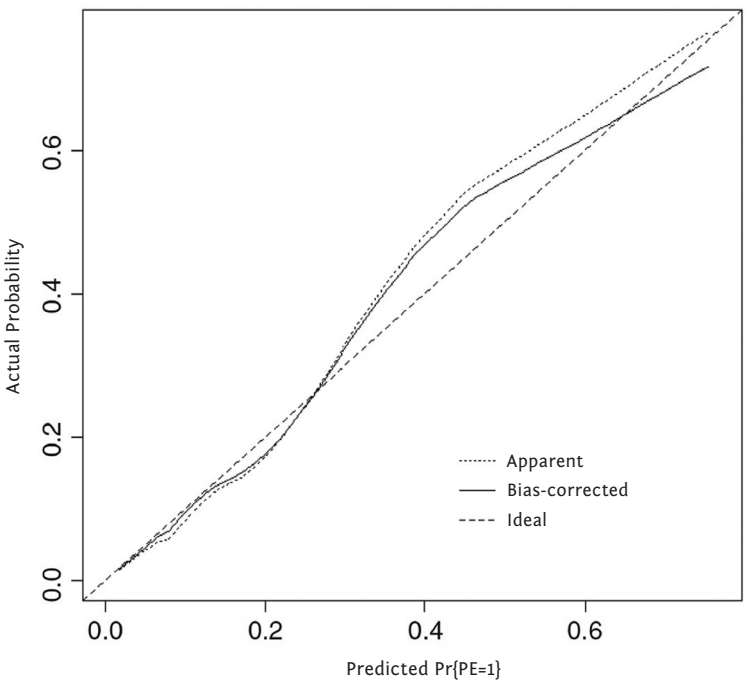
The overall discriminative abilities of both models can be assessed using Receiver Operating Characteristic curves (ROCs). The sensitivities and 1-specificities of both models over all possible probability thresholds are presented in this graph. The higher the areas under this ROCs are, the better the overall discriminative performance of the model with a maximum of 1 and a minimum of 0.5 (diagonal reference line)

INTERNAL VALIDATION

Independent of the approaches used to arrive at the final multivariable model, a major problem in the development phase is the fact that the model has been fitted optimally for the available data. All model development techniques are prone to produce 'overfitted' or overoptimistic and thus unstable models when applied in other individuals, especially if small datasets (limited number of outcomes) or large numbers of predictors are used for model development. (12, 13) To ascertain the best-fitted and most stable model it is essential to continue with a so-called internal validation procedure. (12, 13, 48)

Several techniques are available to evaluate optimism or the amount of overfitting in the developed model. The simplest method is to randomly split the dataset into a development and a validation set and to compare the performance for both models. Since the actual development sample consists of only a part (e.g. 2/3) of the original dataset, which also is not different from the (e.g. 1/3)

FIGURE 2 CALIBRATION CURVE OF MODEL 2 (BASIC MODEL + D-DIMER)



After addition of the D-dimer test to the basic model (see Table 3), the calibration of model 2 was assessed. In the ideal situation, the calibration curve follows the diagonal line in the plot: the predicted probability and observed outcome frequency are the same for all individuals.
Since the model has been fitted optimally in the dataset, there is a chance of overfitting. Therefore, bootstrapping was used to shrink the estimates of the model, indicated by the bias-corrected line.

validation set other than by chance, this method lacks not only efficiency as it decreases statistical power for the model development, it also does not allow for an independent validation of the model. Hence, the random split sample method should preferably not be used. (16, 18, 22)

A more advanced method to avoid waste of development data is the use of bootstrapping. (12, 13, 47) Aim of this procedure is to mimic random sampling from the source population. The sampling procedure consists of multiple samples (e.g. 500) of the same size as the study sample, drawn with replacement (bootstrap). In each sample, all development steps of the model are performed, and indeed different models might be yielded as a result. These bootstrap models are then applied to the original sample. This in turn yields an average estimate of the amount of overfitting or optimism in the originally estimated regression coefficients and predictive accuracy measures, that are adjusted accordingly. (12, 13)

Added value of a new test or biomarker

Often, when developing a prediction model, there is a particular interest in estimating the added – diagnostic or prognostic - predictive value of a new biomarker or (e.g. imaging) test results to existing or established predictors. A good predictive value of such biomarker or test result by itself, i.e. in isolation, is no guarantee for relevant added predictive value when combined with the standard predictors. (64, 66-70) Preferably the new biomarker should be modeled as an extension or supplement to the existing predictors. When developing a diagnostic prediction model following the work-up in practice, this step-by-step and model extension approach is rather sensible: the information of each subsequent test or biomarker result is explicitly added to the previously obtained information, and the target disease probability is adjusted (see *Table 3* (Model 2), *Figure 1* and *2*).

Measures of discrimination such as the AUC (or c-statistic) are insensitive to detecting small improvements in model performance, especially if the AUC of the basic model is already large. (26, 35, 64, 69, 70) Therefore, other measures have been suggested to evaluate the added value of a new biomarker or (imaging) test. Reclassification tables (see *Table 4*) provide insight in the improvement in correct classification of patients. By quantification of the extent to which an extended model improves the correct classification patients into diseased or non-diseased categories, compared to the basic model, the net reclassification improvement (NRI) estimates the added value of the biomarker. Improved classification (a NRI >0.0) suggests that more diseased patients are categorized as high-probability, and non-diseased as low-probability by using the extended model. (69, 71)

The NRI is very dependent on categorization of the probability threshold(s). Different thresholds may result in very different NRIs for the same added test. To overcome this problem of arbitrary cut-off choices, another option is to calculate the so-called integrated discrimination improvement (IDI), which considers the magnitude of the reclassification probability improvement or worsening by a new test over all possible categorizations or probability thresholds. (12, 69, 72)

VALIDATION

If a developed prediction model shows acceptable or good performance based on the internal validation in the development dataset, it is not guaranteed that the model will behave similarly in a different group of individuals. (15, 34)

Therefore, it is essential to assess the performance of the prediction model with patient data not used in the development process and preferably selected by different researchers and in different institutes, countries or even clinical settings or protocols. This is commonly referred to as independent or external validation. (15, 17, 21, 28, 73, 74)

Essentially, formal external validation comprises that in a new set of indivi-

TABLE 4 QUANTIFYING THE ADDED VALUE OF A D-DIMER TEST USING A RECLASSIFICATION TABLE

PE yes (n=73)				
		Model 2 with D-dimer		
		≤25%	**>25%**	**Total**
Model 1	≤25%	22	17	39
without D-dimer	>25%	1	33	34
	Total	23	50	73

PE no (n=525)				
		Model 2 with D-dimer		
		≤25%	**>25%**	**Total**
Model 1	≤25%	445	47	492
without D-dimer	>25%	13	20	33
	Total	458	67	525

Based on the a priori cut-off value chosen, patients are classified as having low or high probability of having PE. In case of high probability, patients are referred to undergo further diagnostic tests (e.g. spiral CT scan).

For this example study, the cut-off value of a 25% PE risk has been chosen arbitrarily for illustration purposes. In clinical practice, a lower cut-off value is obviously be preferred.

In this example, for 17/73 (i.e., 0.23) patients who experienced PE events classification improved with the model with D-dimer, and for 1/73 (0.01) people it became worse, with the net gain in reclassification proportion of 0.22 (thus, 0.23-0.01).

In patients who did not experience an event 47/ 525 (0.09) individuals were reclassified worse by the model with the D-dimer and 13/525 (0.02) were reclassified better, resulting in a net gain in reclassification proportion of -0.07.

The total net gain in reclassification proportion (NRI) therefore was 0.22- 0.07 = 0.15 (95% CI 0.04–0.27).

The integrated discrimination improvement (IDI) was: 0.09 (95% CI 0.08-0.11)

duals, the predicted outcome probabilities are estimated by using the originally developed model, and compared to the actual outcomes. Importantly, external validation is not repeating the analytic steps or refitting the developed model in the new validation data and then comparing the model performance. (15, 17, 22, 74) Three methods of external validation are available and can be carried out in a prospective manner, but also retrospectively if datasets with the necessary information on predictors and outcomes are available. (15, 17, 22, 28, 34, 73, 74)

TEMPORAL VALIDATION

Temporal validation may be done by splitting a large (development) dataset non-randomly based on the moment of participant inclusion. (15, 17, 18, 22) One may argue this is not a form of independent or external validation but a form of non-random split-sample internal validation, as the entire data set is established by

the same researchers using the same definitions and measurements. However, it results in more variation between the development and validation sample than random splitting. (17) Prospective evaluation of the model in a new study sample by the same researchers, in the same institutions only later in time might allow for more variation. (17) For example, to develop a DVT prediction model for a primary care setting, Oudega et al. studied a large prospective cohort of suspected patients. (11) The newly developed rule was then validated in largely the same primary care practices but with participants recruited during a later time period, by Toll et al. (75)

GEOGRAPHICAL VALIDATION

As with temporal validation, one may assess the performance of a prediction model in other institutes or countries, by non-randomly splitting a large development dataset based on institute or country. (17) A more external or independent validation is when the model is validated in other institutes or country by different researchers, as has been done by Klok and colleagues for the Revised Geneva Score to diagnose PE. (76) Due to more variation in case-mix (in- and exclusion criteria chosen) and even in measurements of predictors and outcome, the latter variant provides a more thorough and independent validation. Obviously, external validations may include a combination of temporal and geographical validation.

DOMAIN VALIDATION

Perhaps the most extreme and rigid form of external validation is the assessment of the prediction model in a completely different clinical domain or setting. (15, 17, 22, 28, 34, 73, 74) Domain validation may e.g. comprise a model developed in secondary care and validated in primary care, developed in adults and validated in children or developed for predicting fatal events and validated for its ability to predict non-fatal events.

Model updating or adjustment

The external validation procedure provides quantitative information on the discrimination, calibration and classification of the model in a population that differs from the development population. (15, 22, 28, 73, 74) Ideally, the performance is comparable in the development and validation sample, indicating that the model can be used in the source populations of both. (15) But often the model performance in the new individuals is worse than that found in the development study. This does not mean that the model should be refrained from further using. Whereas it might be tempting to just built a new model that will, obviously, fit the data at hand more properly, this increases the wild growth of models for the same clinical situation. (4, 12, 17, 22, 77-79)

There are no strict criteria how to define poor or acceptable performance. (28, 58, 73, 74) If prediction model performance is considered to perform poorly, the original model can be adapted to the circumstances of the validation sample. (22, 77-79) These so-called updating methods include very simple adjustment of the

baseline risk, simple adjustment of predictor weights, re-estimation of predictors weights, or addition or removal of predictors, and have been described extensively elsewhere. (12, 34, 77-80) The updated prediction model should preferably be externally validated as well. (4, 17)

From a clinical perspective, external validation is often approached differently. Instead of pursuing the most optimal fit of a model, the main question is whether patient outcomes, e.g failure rate and incorrect predictions, remain acceptable and adequate if the model is applied in another population. For example, the AMUSE-2 study validated the use of the Wells PE rule in a primary care setting by comparing its efficiency (i.e. the proportion of patients categorized as low risk by the Wells PE rule and D-dimer testing) and safety (i.e. the proportion of "missed" PE cases in this low risk group) with generally accepted failure rates from secondary care studies. (32) Further updating was not considered.

IMPACT AND IMPLEMENTATION

The final step towards implementation of a developed and validated (and if needed updated) prediction model is the quantification of the impact when it is actually used to direct patient management in clinical care. (4, 17, 22, 28, 74) To what extent contributes the use of the prediction model to the (change in) behavior and (self-) management of patients and doctors? And ultimately, what are the effects on health outcomes and cost-effectiveness of care? Although this final step is important to improve health care, reviews showed that this form of prediction modeling studies is even less frequently performed than external validation studies. (22, 28, 29)

The largest difference from a validation study is the fact that impact studies require a control group. (4, 17, 28) It is essential to compare the effects on decision-making and health outcomes by using standard care or by prediction model guided care.

A major disadvantage of the ordinary RCT design - in which each consecutive patient can be randomized to either the index (prediction model guided management) or control (care-as-usual) - is the impossibility of blinding and subsequently the potential learning curve of the treating physicians. This makes eventually the two groups increasingly alike and dilutes the potential effect. (4, 17) These learning effects are prevented by randomization of clusters rather than patients. All patients within a cluster, for example a doctor or hospital, receive the same type of intervention. (81) Unfortunately, cluster RCTs do require more individuals to obtain the same amount of power, compared to the standard RCT design, and are therefore often costly to perform.

The stepped wedge design is an appealing variant of the standard cluster-randomized trial if the new, often complex intervention has to be implemented in routine care. (17, 82) All clusters (e.g. hospital) will switch from usual care to the intervention eventually, but the exact moment of transition is randomly assigned across the clusters. This design improves the statistical efficiency. Moreover, potential problems in implementation of the new intervention can be detected

early in the course of the trial and thus reacted upon immediately.

There are also several non-randomized study designs that can be used to assess impact, and might even be worthwhile to conduct before deciding to start a cluster (stepped wedge) RCT. (4, 17) A prospective before-after impact study compares patient outcomes before and after implementation of the prediction model. Although less complex and time-consuming, it is prone to potential time effects and subject differences. A before-after study within the same doctors is even simpler. Doctors are asked to document the treatment decision before and after exposure to the prediction model for the same patient. No follow-up is involved and it is easy and cheap to perform, but as with the prospective before-after studies, there is the potential of time effects. An interesting and cost-efficient alternative for impact studies that would require a long follow-up period is the performance of a decision analytic model. (17, 68, 83-85) This is a mathematical approach to combine information on patient outcomes and health effects, usually from prior RCTs or meta-analyses, with the predicted accuracy measures and their uncertainty as presented by the prediction model. These models estimate the (cost-) effectiveness of implementation of the prediction model in clinical daily care, as compared to usual care. If the outcomes show that the new prediction model does not improve clinical care and thus patient outcomes, one might wonder if a (often costly and time-consuming) trial is worthwhile to be performed. (17, 68) As an example of a decision analytic model, we refer to the cost-effectiveness analysis of using of a primary care clinical decision rule, combined with a qualitative D-dimer test, in suspected DVT. (86)

CONCLUDING REMARKS

The advantages of using risk prediction models in clinical care – namely more individually risk tailored management and thus increase in efficiency and ultimately cost-effectiveness – drive the popularity of developing and using prediction models. Yet, despite this popularity, there is also concern that the use of prediction models will lead to so-called "cookbook medicine", a situation in which the doctor's gut-feeling (or gestalt) is completely bypassed by the use of prediction rules. (14, 28, 87) We believe that that probabilities estimated by a prediction model are not considered to replace but rather help the doctor's decision making. (4, 14, 17) In fact, it should serve as a useful tool to incorporate all the single pieces of information to aid their clinical reasoning. Sometimes, physicians are as good as prediction models to identify those individuals that actually are diseased whereas prediction models are better or more efficient in identifying those individuals where a disease can be excluded. For instance, a recent meta-analysis by Lucassen and colleagues showed that gestalt (that is, the estimated probability of a patient being diseased or not based on clinical reasoning) is just as sensitive as the

application of a risk prediction model to rule out PE, but much less specific. (39) As a consequence, although no more PE cases are actually missed by physicians using their own gut feeling yet many more patients are unnecessarily referred for spiral CT scanning. Of note, this not only is associated with higher costs but also poses more patients with the inherent risks of CT scanning: radiation and contrast nephropathy. For an ultimate answer on the cost-effectiveness of the use and thus impact of this model, a decision analytic model should still be performed, as discussed above.

To conclude, we aimed to provide a comprehensive overview of the steps in risk prediction modeling – from development to validation to impact assessment - the preferred methodology per step, and the potential pitfalls to overcome. We hope this will guide future research on this topic, and enhance applied studies of risk prediction modeling in the field of thrombosis and hemostasis.

BOX 1 EXAMPLES OF RISK PREDICTION MODELS IN THE VTE DOMAIN

	First Author	Clinical Context
Derivation		
Oudega PM for DVT	R. Oudega	The Wells DVT rule was not safe in primary care, and therefore a new prediction model for primary care was developed.
Vienna Prediction Model	S. Eichinger	VTE recurrence risk is high in patients with a first (unprovoked) event, yet is actual risk in individual patients is unknown.
PESI model	D. Aujesky	Outpatient treatment of patients with PE may be safe; the PESI model was developed to identify patients with a low risk of short-term mortality in whom this indeed may be safe.
Validation		
AMUSE-2 study	G.J. Geersing	This study validated the Wells PE rule as a safe tool in patients suspected of PE in a primary care domain.
Revised Geneva Score for PE	F.A. Klok	The Revised Geneva Score for PE was validated in new cohort of patients.
Validation study Oudega PM for DVT	D.B. Toll	This study validated the Oudega rule for DVT for different subgroups, i.e. based on age, gender, and previous VTE.
Impact		
Wells DVT rule with D-dimer	P. Wells	In this landmark RCT the safety of not performing CUS in patients with a low Wells score and a negative D-dimer test was demonstrated.
CEA study for AMUSE-1 strategy	A. Ten Cate-Hoek	In this study, all costs and effects of not referring a patient with suspected DVT and a low score on the Oudega rule were quantified, demonstrating its cost-effectiveness
OTPE trial for outpatient PE management	D. Aujesky	This RCT demonstrated that it is safe to treat patients out of the hospital if their PESI score is low (PESI classes I and II).

PM = prediction model; DVT = deep venous thrombosis; PE = pulmonary embolism; VTE = venous thromboembolism; PESI = Pulmonary Embolism Severity Index; CEA = cost-effectiveness analysis; $ using backwards stepwise selection; # clinical variables in the model were selected based on significance levels, for laboratory variables forwards selection; ± selection process (backwards or forwards) not described in detail in paper; ^ most diagnostic validation studies on the validation of PMs in suspected VTE do not routinely evaluate model validity in a formal manner, yet focus on clinical outcomes only, i.e. safety = proportion of "missed" VTE cases in a low-risk population, and efficiency = proportion of patients identified as low-risk (see main text for more detail on the difference between formal model

	N	Type of Model	Outcome(s)	Ref
	1,295	Logistic regression $	Confirmed DVT	(11)
	929	Cox regression #	Recurrence of VTE during a median FU of 43.3 months.	(8)
	10,354	Logistic regression. ±	Death from any cause within 30 days.	(10)
	598	Safety and efficiency ^	Confirmed PE	(32)
	300	c-statistic	Confirmed PE	(76)
	2,086	Safety and efficiency ^	Confirmed DVT	(75)
	601 ¥	Safety and efficiency ^	Confirmed DVT	(88)
	N.A.	Markov model	Costs per QALY and iCER	(86)
	344	Proportion of outcome in index and control group	Symptomatic recurrent DVT or PE within 90 days.	(89)

validity analyses and clinical outcome validity); ¥ the study randomized 601 patients with a low score on the Wells PM for DVT (score ≤ 1) into a D-dimer testing group and no D-dimer testing group; in addition it also randomized 495 DVT likely patients, yet these data are less relevant when the aim is to exclude DVT; N.A. = not applicable as this is cost-effectiveness modeling study; QALY = quality-adjusted life year; iCER = incremental cost-effectiveness ratio

REFERENCES

1. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. N Engl J Med. 1985;313(13):793-9.
2. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277(6):488-94.
3. Grobbee DE, Hoes AW. Clinical epidemiology- Principles, methods and applications for clinical research. London: Jones and Bartlett Publishers; 2009.
4. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
5. Wells PS, Anderson DR, Bormanis J, Guy F, Mitchell M, Gray L, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. Lancet. 1997;350(9094):1795-8.
6. Wells P, Andersen D, Rodger M, Ginsberg JS, Kearon C, Gent M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. Thromb Haemost. 2000;83(3):416-20.
7. Rodger M, Kahn S, Wells P, Anderson D, Chagnon I, Le Gal G, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. CMAJ. 2008;179(5):417-26.
8. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation. 2010;121(14):1630-6.
9. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). J Thromb Haemost. 2012;10(6):1019-25.
10. Aujesky D, Obrosky DS, Stone RA, Auble TE, Perrier A, Cornuz J, et al. Derivation and validation of a prognostic model for pulmonary embolism. Am J Respir Crit Care Med. 2005;172(8):1041-6.
11. Oudega R, Moons KG, Hoes AW. Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. Thromb Haemost. 2005;94(1):200-5.
12. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2008.
13. Harrell FE. Regression modeling strategies. New York: Springer-Verlag; 2001.
14. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
15. Altman D, Vergouwe Y, Royston P, Moons K. Prognosis and prognostic research: validating a prognostic model. Bmj. 2009;338:b605.
16. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338:b604.
17. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.

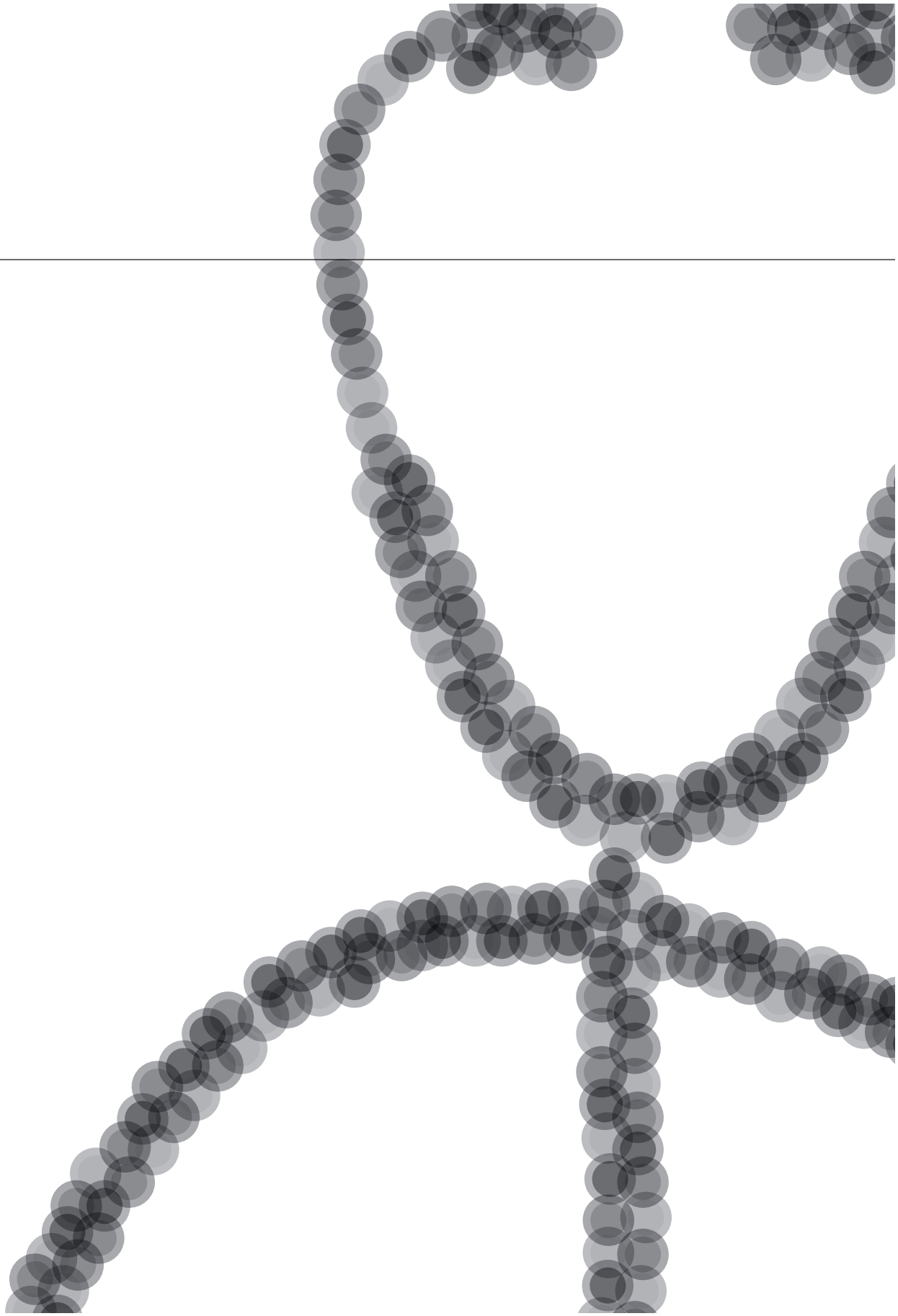
18. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683-90.
19. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ. 2013;346:e5595.
20. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ. 2013;346:e5793.
21. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic factor research. PLoS Med. 2013;10(2):e1001380.
22. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med. 2013;10(2):e1001381.
23. Bouwmeester W, Twisk JW, Kappen TH, Klei WA, Moons KG, Vergouwe Y. Prediction models for clustered data: comparison of a random intercept and standard regression model. BMC Med Res Methodol. 2013;13(1):19.
24. Mallett S, Royston P, Dutton S, Waters R, Altman D. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010;8(1):20.
25. Mallett S, Royston P, Waters R, Dutton S, Altman D. Reporting performance of prognostic models in cancer: a review. BMC Med. 2010;8(1):21.
26. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. JAMA. 2009;302(21):2345-52.
27. Collins G, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med. 2011;9(1):103.
28. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-9.
29. van Dieren S, Beulens JWJ, Kengne AP, Peelen LM, Rutten GEHM, Woodward M, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart. 2012;98(5):360-9.
30. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):e1001221.
31. Moons KG. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem. 2010;56(4):537-41.
32. Geersing GJ, Erkens PMG, Lucassen W, Buller H, Ten Cate H, Hoes A, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
33. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
34. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085-94.
35. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. Clin Chem. 2012;58(10):1408-17.

36. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50(3):473-6.
37. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. A basic science for clinical medicine. 2nd ed. ed. Boston: Little, Brown and Company; 1991.
38. Moons KG, Grobbee DE. Diagnostic studies as multivariable, prediction research. J Epidemiol Community Health. 2002;56(5):337-8.
39. Lucassen W, Geersing G, Erkens PMG, Reitsma JB, Moons KGM, Büller H, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. Ann Intern Med. 2011;155(7):448-60.
40. Kearon C, Iorio A, Palareti G, on behalf of the subcommittee on control of anticoagulation of the SSC of the I. Risk of recurrent venous thromboembolism after stopping treatment in cohort studies: recommendation for acceptable rates and standardized reporting. J Thromb Haemost. 2010;8(10):2313-5.
41. Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. BMC Med Res Methodol. 2008;8:48.
42. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005;51(8):1335-41.
43. Ganna A, Reilly M, de Faire U, Pedersen N, Magnusson P, Ingelsson E. Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease. Am J Epidemiol. 2012;175(7):715-24.
44. Altman D, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332(7549):1080.
45. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127-41.
46. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med. 2007;26(30):5512-28.
47. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059-79.
48. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49(8):907-16.
49. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503-10.
50. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. J Clin Epidemiol. 1995;48(12):1495-501.
51. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
52. Janssen KJ, Donders AR, Harrell FE, Jr., Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol. 2010;63(7):721-7.

53. Janssen KJ, Vergouwe Y, Donders AR, Harrell FE, Jr., Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. Clin Chem. 2009;55(5):994-1001.
54. Vach W. Some issues in estimating the effect of prognostic factors from incomplete covariate data. Stat Med. 1997;16(1-3):57-72.
55. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol. 2006;59(10):1102-9.
56. Gorelick MH. Bias arising from missing data in predictive models. J Clin Epidemiol. 2006;59(10):1115-23.
57. Wood A, White I, Royston P. How should variable selection be performed with multiply imputed data? Stat Med 2008;27(17):3227-46.
58. Vergouwe Y, Royston P, Moons K, Altman D. Development and validation of a prediction model with missing predictor data: a practical approach. J Clin Epidemiol. 2009;63:205 - 14.
59. de Groot JA, Janssen KJ, Zwinderman AH, Moons KG, Reitsma JB. Multiple imputation to correct for partial verification bias revisited. Stat Med. 2008;27(28):5880-9.
60. Little RA. Regression with missing X's; a review. J Am Stat Assoc. 1992;87:1227-37.
61. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087-91.
62. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361 - 87.
63. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
64. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928-35.
65. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. ed. New York: Wiley; 2000.
66. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology. 1997;8(1):12-7.
67. Moons KG, van Es GA, Michel BC, Buller HR, Habbema JD, Grobbee DE. Redundancy of single diagnostic test evaluation. Epidemiology. 1999;10(3):276-81.
68. Moons KGM. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem. 2010;56(4):537-41.
69. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157-72.
70. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, et al. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008;167(3):362-8.
71. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med. 2009;150(11):795-802.
72. Pencina MJ, D'Agostino RB, Vasan RS. Statistical methods for assessment of added usefulness of new biomarkers. Clin Chem Lab Med. 2010;48(12):1703-11.

73. Justice A, Covinsky K, Berlin J. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130:515 - 24.
74. Altman D, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19:453 - 73.
75. Toll DB, Oudega R, Vergouwe Y, Moons KG, Hoes AW. A new diagnostic rule for deep vein thrombosis: safety and efficiency in clinically relevant subgroups. Fam Pract. 2008;25(1):3-8.
76. Klok FA, Kruisman E, Spaan J, Nijkeuter M, Righini M, Aujesky D, et al. Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism. J Thromb Haemost. 2008;6(1):40-4.
77. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.
78. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61(1):76-86.
79. Janssen KJ, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KG. A simple method to adjust clinical prediction models to local circumstances. Can J Anaesth. 2009;56(3):194-201.
80. van Houwelingen H. Validation, calibration, revision and combination of prognostic survival models. Stat Med. 2000;19:3401 - 15.
81. Campbell M, Elbourne D, Altman D. CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328(7441):702-8.
82. Brown C, Lilford R. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006;6(1):54.
83. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Basic types of economic evaluation. Methods for the economic evaluation of health care programmes. . 3rd ed. New York: Oxford University Press; 2005.
84. Schaafsma JD, van der Graaf Y, Rinkel GJE, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. J Clin Epidemiol. 2009;62(12):1248.
85. den Ruijter HM, Vaartjes I, Sutton-Tyrrell K, Bots ML, Koffijberg H. Long-term health benefits and costs of measurement of carotid intima–media thickness in prevention of coronary heart disease. J Hypertens. 2013;31(4):782-90.
86. Ten Cate-Hoek AJ, Toll DB, Buller HR, Hoes AW, Moons KG, Oudega R, et al. Cost-effectiveness of ruling out deep venous thrombosis in primary care versus care as usual. J Thromb Haemost. 2009;7(12):2042-9.
87. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.
88. Wells PS, Anderson DR, Rodger M, Forgie M, Kearon C, Dreyer J, et al. Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. N Engl J Med. 2003;349(13):1227-35.
89. Aujesky D, Roy P-M, Verschuren F, Righini M, Osterwalder J, Egloff M, et al. Outpatient versus inpatient treatment for patients with acute pulmonary embolism: an international,

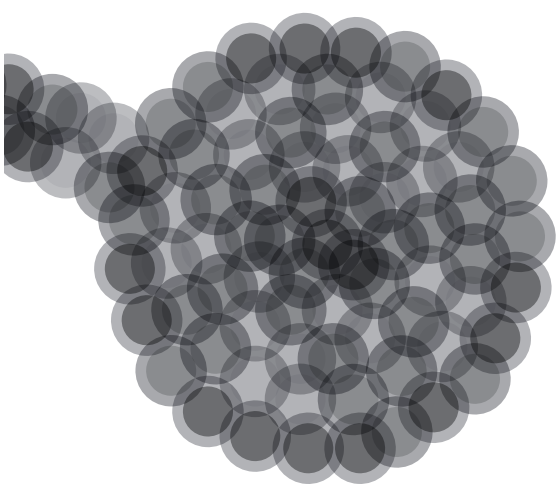
open-label, randomised, non-inferiority trial. Lancet. 2011;378(9785):41-8.
90. Ten Cate-Hoek AJ, Prins MH. Management studies using a combination of D-dimer test result and clinical probability to rule out venous thromboembolism: a systematic review. J Thromb Haemost. 2005;3(11):2465.



CHAPTER 3

DIAGNOSTIC PREDICTION MODELS FOR SUSPECTED PULMONARY EMBOLISM: SYSTEMATIC REVIEW AND INDEPENDENT EXTERNAL VALIDATION IN PRIMARY CARE

Janneke MT Hendriksen, Geert-Jan Geersing, Wim AM Lucassen,
Petra MG Erkens, Henri EJH Stoffers, Henk CPM van Weert,
Harry R Büller, Arno W Hoes, Karel GM Moons



BMJ 2015;351:h4438.

ABSTRACT

Objective: To validate all diagnostic prediction models for ruling out pulmonary embolism that are easily applicable in primary care.
Design: Systematic review followed by independent external validation study to assess transportability of retrieved models to primary care medicine.
Setting: 300 general practices in the Netherlands.
Participants: Individual patient dataset of 598 patients with suspected acute pulmonary embolism in primary care.
Main outcome measures: Discriminative ability of all models retrieved by systematic literature search, assessed by calculation and comparison of C statistics; sensitivity, specificity, efficiency (overall proportion of patients with low probability of pulmonary embolism), and failure rate (proportion of pulmonary embolism cases in group of patients with low probability) for all models, with stratification into groups with high and low probability of pulmonary embolism according to pre-specified model cut-offs combined with qualitative D-dimer test.
Results: Ten prediction models for the diagnosis of pulmonary embolism have been published. Five of these models could be validated in the primary care dataset: the original Wells, modified Wells, simplified Wells, revised Geneva, and simplified revised Geneva models. Discriminative ability was comparable for all models (range of C statistic 0.75-0.80). Sensitivity ranged from 88% (simplified revised Geneva) to 96% (simplified Wells) and specificity from 48% (revised Geneva) to 53% (simplified revised Geneva). Whereas efficiency of all models was between 43% and 48%. Differences were observed between failure rates, especially between the simplified Wells and the simplified revised Geneva models (failure rates 1.2% (95% confidence interval 0.2% to 3.3%) and 3.1% (1.4% to 5.9%), respectively; absolute difference –1.98% (–3.33% to –0.74%)). Irrespective of the diagnostic prediction model used, three patients were incorrectly classified as having low probability of pulmonary embolism; pulmonary embolism was diagnosed only after referral to secondary care.
Conclusions: Five diagnostic pulmonary embolism prediction models that are easily applicable in primary care were validated in this setting. Whereas efficiency was comparable for all rules, the Wells rules gave the best performance in terms of lower failure rates.

INTRODUCTION

Pulmonary embolism is a potentially fatal condition if left untreated. Its presentation can be relatively mild, sometimes even mimicking myalgia or a simple cough. This causes pulmonary embolism to be a diagnosis easily missed. (1, 2) As a result, physicians have a low threshold for suspicion and subsequent referral for further diagnostics. (3, 4) Referred patients will be exposed to the burden, costs, and even potential iatrogenic damage of diagnostic techniques such as spiral computed tomography or contrast nephropathy. (5) However, only in a small subset (about 10-15%) of all suspected cases are emboli actually confirmed during diagnostic investigation. (6)

Several non-invasive diagnostic prediction models have been developed for safe exclusion of pulmonary embolism and are usually followed by D-dimer testing. (7) Physicians can use these models as a strategy to enhance the efficiency of the diagnostic process by precluding those patients with a low probability of pulmonary embolism from further diagnostic tests, without compromising on safety (that is, missing cases of pulmonary embolism). Such diagnostic strategies can reduce the number of unnecessary computed tomography scans by 35%, with only 1-2% of missed cases in the group of patients with a low probability of pulmonary embolism. (7)

In many countries, general practitioners are the first physicians to encounter patients with symptoms suggestive of pulmonary embolism. As the decision must be made whom to refer or not, risk stratification is particularly valuable. All diagnostic models for safe exclusion of pulmonary embolism have been developed and validated in hospital or acute care settings. However, diagnostic prediction models developed in a particular setting often perform less well when applied in another setting. Therefore, models derived in hospital or acute care settings cannot simply be implemented in primary care. (8-14) Reasons for this poorer performance include differences in the case mix and the prevalence of pulmonary embolism due to the unselected population, as well as differences in physicians' experience of patients with suspected pulmonary embolism. (9, 10, 15, 16) Hence, when transferring diagnostic models or strategies across healthcare settings, evaluation of their performance in this other setting is necessary first. This form of external validation is referred to as domain or setting validation, (8, 10, 17) or as quantification of the transportability of prediction models. (13, 18)

The recent AMUSE-2 study (Amsterdam, Maastricht, Utrecht Study on thrombo-Embolism) (19) has been the first to prospectively quantify the transportability of the, perhaps best known, secondary care derived diagnostic prediction model for pulmonary embolism (that is, the Wells pulmonary embolism rule, (20) combined with point of care D-dimer testing) in a primary care setting. Various other diagnostic pulmonary embolism prediction models that may also be valuable for primary care have been developed but have not been validated in a primary care population.

The aim of this study was therefore to assess the clinical performance in a primary care setting of all existing diagnostic models developed for patients with suspected pulmonary embolism. We firstly did a systematic review and critical appraisal of all available diagnostic models for pulmonary embolism, as recommended by guidelines on prediction models research. (21) Next, the diagnostic models easily applicable in primary care were validated in the AMUSE-2 dataset—that is, a large independent prospectively constructed cohort of patients presenting to their general practitioner with complaints suggestive of pulmonary embolism.

METHODS

UPDATED SYSTEMATIC REVIEW

For our systematic review and critical appraisal of the existing diagnostic models for pulmonary embolism, we followed the recent methodological guidance by the Prognosis Methods Group of the Cochrane Collaboration. (21-24)

Firstly, we framed the review question and design by using the CHARMS checklist for systematic reviews of prediction models (see Appendix Box A). (21) We then repeated the systematic search previously performed for an aggregate meta-analysis by Lucassen et al and used the same study selection criteria. (7) We searched for studies on development and validation of diagnostic prediction models published between January 2010 and October 2014. Details on the search syntax can be found in Appendix *Figure A*. We then critically appraised studies on the development of diagnostic prediction models by using the CHARMS checklist (see Appendix *Table A*). All retrieved papers were examined by two independent reviewers (JH, GJG) and a third independent reviewer (KGMM) in case of disagreement.

Given the scope of our systematic review (see Appendix Box A), we assessed all diagnostic prediction models for pulmonary embolism, retrieved by our search, on their applicability in a primary care domain. Accordingly, the diagnostic predictors or tests included in the diagnostic model needed to be measurable at the general practitioner's office. Variables such as signs and symptoms, items from history taking, pulse rate, or blood pressure are easily and quickly obtained in primary care, whereas results from (arterial) blood gas analyses, chest radiographs, or advanced electrocardiograph interpretations generally are not. Diagnostic models with predictors that cannot easily be obtained in primary care were excluded from the main analyses.

VALIDATION COHORT

Population characteristics

The AMUSE-2 cohort was designed to prospectively validate the Wells pulmonary embolism rule in a Dutch primary care setting. The study took place between 1 July 2007 and 31 December 2010. In short, it included 662 adult patients presenting

at one of the participating general practices with complaints raising suspicion of pulmonary embolism (that is, acute dyspnoea, pain on inspiration, or unexplained cough; all at the discretion of the including physicians). Of these patients, 64 met one of the predefined exclusion criteria of anticoagulant treatment at presentation, pregnancy, or unwillingness or inability to provide written informed consent, leaving 598 patients for further evaluation. More details of this cohort and the sample size calculation are described elsewhere. (19)

Predictors

In all participants, the general practitioner assessed relevant information on general health and specific cardiopulmonary and signs and symptoms of deep venous thrombosis by systematically filling out a pre-specified case record form. Subsequently, a qualitative point of care D-dimer test (Simplify D-dimer; Clearview, Inverness Medical, Bedford, UK) was performed. This test returns a visual dichotomous outcome; a positive test result is indicated by a pink-purple coloured line that appears on the disposable device within 10 minutes of application. This corresponds to a D-dimer concentration above 80 ng/mL. Only a control line will be visible if the test is negative. In case of an inconclusive test result, we classified the result as positive.

All predictors in the validation cohort were assessed blinded for the outcome. Exact definitions and measurement methods of the predictors in the validation cohort have been described previously. (19)

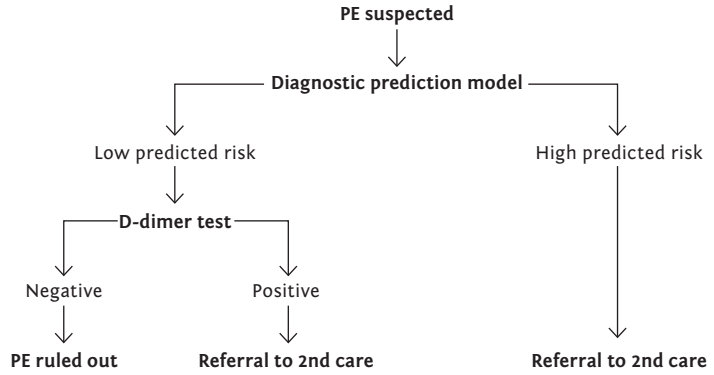
Outcome

The study protocol recommended referral of all patients with suspected pulmonary embolism to secondary care, regardless of the outcome of the Wells rule or D-dimer test. In secondary care, the regular diagnostic pathway according to local hospital guidelines was followed, with no explicit blinding for the general practitioner's findings. This usually comprised a combination of estimated probability and quantitative laboratory based D-dimer testing and was followed by diagnostic imaging if indicated. The primary outcome was the presence of venous thromboembolism (either deep venous thrombosis or pulmonary embolism), as based on a composite reference standard of all diagnostic imaging tests performed in the hospital (spiral computed tomography, ventilation-perfusion scanning, pulmonary angiography, leg ultrasonography, and clinical probability assessment as performed in secondary care, with or without D-dimer testing) and including any occurrence of venous thromboembolic events during three months of follow-up in primary care.

DATA ANALYSIS

For all diagnostic prediction models, we retrospectively calculated the individual score of each included patient on the basis of the presence or absence of the model's predictors. We compared the overall discriminative ability of the models,

FIGURE 1 FLOW SCHEME OF DIAGNOSTIC PATHWAY IN SUSPECTED PULMONARY EMBOLISM IN PRIMARY CARE



PE= pulmonary embolism.

using the c-statistic (that is, the area under the curve) of the receiver operating characteristics curve, with 95% confidence intervals. We assessed differences between the c-statistics with the DeLong method. (25)

To stratify all participants in the validation cohort into categories of low or high probability of having pulmonary embolism, we used the diagnostic pathway as recommended by guidelines (see *Figure 1*): first the stratification based on the cut-off value of each diagnostic model as suggested in the development papers for the model, followed by a D-dimer test in case of a low predicted probability of pulmonary embolism. A subsequent negative D-dimer test implied a low predicted probability of pulmonary embolism and no need for referral to secondary care. In all other cases, the high predicted probability of pulmonary embolism meant referral for further objective testing.

Given the prevalence in this primary care cohort and following previous publications on this topic, we chose to use the first of these cut-off values for all subsequent analyses. (19, 26)

We then calculated the common diagnostic accuracy measures of the models (that is, sensitivity, specificity, and positive and negative predictive values).

From a clinical point of view, a diagnostic prediction model should ideally classify as many patients as possible in the non-referral group, but not at the expense of an increase in pulmonary embolism events missed in this group. Therefore, we evaluated the clinical performance of each model (combined with D-dimer) by focusing on the efficiency and the failure rate. We defined efficiency as the proportion of patients in the whole cohort stratified to the group with low

predicted probability of pulmonary embolism (that is: (true negatives (tn)+false negatives (fn))/ total cohort). We defined failure rate as the proportion of these patients with low predicted probability of pulmonary embolism ultimately diagnosed as having pulmonary embolism on the basis of our composite reference standard (that is: fn/ (tn+fn)). We calculated differences in failure rates between the models, with the surrounding 95% confidence intervals. We then varied the cut-off values as proposed in the model development studies to evaluate the influence of different cut-off values on the outcome measures.

Finally, we constructed calibration plots for the diagnostic prediction models. With calibration plots, the agreement between the predicted and observed probability of pulmonary embolism can be visualised. In the absence of a reported intercept for the models, we re-estimated the intercept in the validation cohort by using the linear predictor as offset in a logistic regression model including the model coefficients.

In the dataset, we imputed missing values for predictors by using multiple imputation techniques before our analyses. (27, 28) Imputation was performed to minimise the effect of the bias associated with selectively ignoring these patients. In case of a high percentage of missing values, no imputation was performed. Instead, the main analysis was carried out with all missing values assigned as the predictor being absent. In a sensitivity analysis, all missing values were assigned as the predictor being present.

We used IBM SPSS version 21 for descriptive statistical analyses. We used R version 3.2.0 for forest plots and the calculation of differences in c-statistics (DeLong method) and failure rates.

REPORTING

The results of this validation study were reported in adherence to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guideline (see Appendix *Table B*). (29, 30)

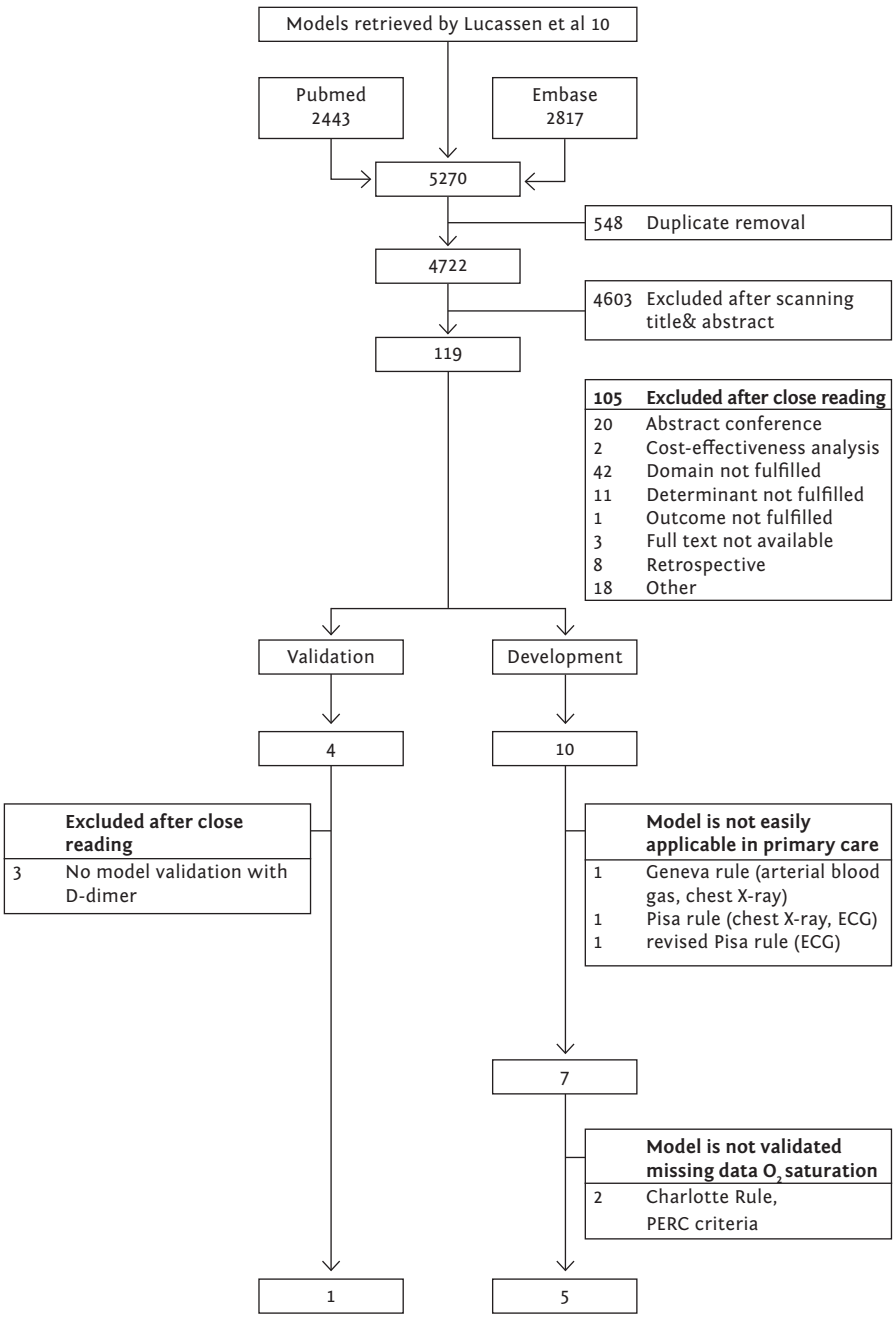
PATIENT INVOLVEMENT

No patients were involved in setting the research question or the outcome measures; nor were they involved in recruitment or in the design and implementation of the study. There are no plans to involve patients in dissemination.

RESULTS

Our systematic literature search identified four model validation studies, but no newly developed models in addition to the 10 models previously found with the search by Lucassen (see *Figure 2*). Of these 10 models, we excluded two models (the Pisa rule (31) and the original Geneva score (32)) from further analysis on the basis of the predetermined criteria for validation in our primary care cohort (see

FIGURE 2 OVERVIEW OF SELECTION OF STUDIES THAT DEVELOPED OR VALIDATED PREDICTION MODELS FOR DIAGNOSIS OF PULMONARY EMBOLISM, BASED ON LITERATURE SEARCH IN PUBMED AND EMBASE



PERC=pulmonary embolism rule-out criteria. ECG= electrocardiography.

FIGURE 3 FOREST PLOT OF FAILURE RATES IN DEVELOPMENT AND VALIDATION STUDIES OF DIAGNOSTIC PREDICTION MODELS, IF COMBINED WITH D-DIMER TESTING

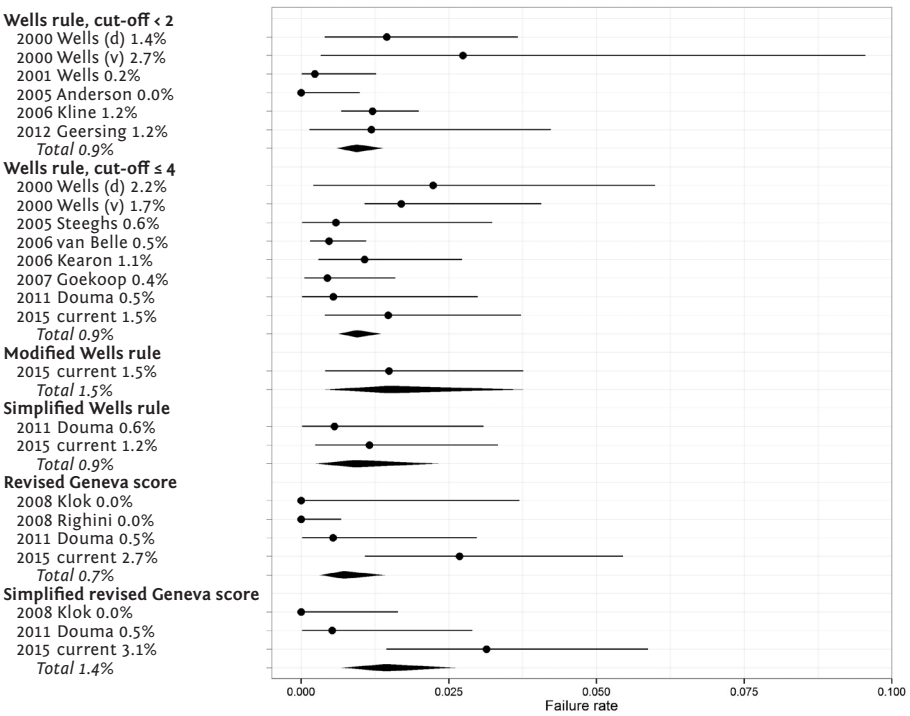


Figure 2). The original, modified, and simplified Wells rules, the revised Geneva scores, the revised Pisa rule, the pulmonary embolism rule-out criteria (PERC) rule, and the Charlotte rule fulfilled the predetermined primary care criteria. (20, 33-38) However, owing to complete missing information on the predictors oxygen saturation and body temperature in our dataset, we were not able to validate the revised Pisa rule, the PERC rule, and the Charlotte rule, leaving five models for further evaluation. See Appendix *Table C* for an overview of all diagnostic models identified in this study, including the five models that were tested in the individual patient dataset of our validation cohort: the original, modified, and simplified Wells rules and the original and simplified revised Geneva scores.

Of the four recently published validation studies identified by our search, only one study reported the results of the combination of the prediction model and D-dimer testing and was used in this current update. (19) The observed failure rates from previous development and validation studies, and the current validation of the prediction models, are shown in a forest plot (see *Figure 3*).

TABLE 1 BASELINE CHARACTERISTICS OF PARTICIPANTS IN EXTERNAL VALIDATION COHORT (AMUSE-2; PATIENTS PRESENTING IN PRIMARY CARE WITH SUSPECTED PULMONARY EMBOLISM), AND BASELINE CHARACTERISTICS OF DEVELOPMENT STUDIES OF FIVE DIAGNOSTIC PREDICTION MODELS. VALUES ARE NUMBERS (PERCENTAGES) UNLESS STATED OTHERWISE

Characteristics	**AMUSE-2 validation cohort** (n=598)	**Original Wells (20)** (n=1260)
Male sex	173 (28.9)	NR
Mean (SD) age, years	48 (16)	NR
Venous thromboembolic events	73 (12.2)	NA
Pulmonary embolism cases	72 (12.0)	222 (17.6)
Items Wells rule		
Clinical signs and symptoms of DVT	57 (9.5)	NR
Alternative diagnosis less likely	333 (55.7)	NR
Heart rate >100 bpm	111 (18.6)	NR
Immobilisation or recent surgery	94 (15.7)	NR
Previous pulmonary embolism or DVT	84 (14.0)	160 (12.7)
Haemoptysis	21 (3.5)	NR
Malignancy	26 (4.3)	NR
Items revised Geneva score		
Age >65 years	103 (17.2)	NR
Heart rate:		
<75 bpm	187 (31.3)	NR
75-94 bpm	259 (43.3)	NR
>94 bpm	152 (25.4)	NR
Lower limb pain	96 (16.1)	NR
Oedema/ swelling	48 (8.0)	NR
Pain and swelling*†	22 (3.7)	NR
Pain on lower limb deep venous palpation†	34 (5.7)	NR
Pain and swelling*‡	24 (4.0)	NR
Pain on lower limb deep venous palpation‡	391 (65.4)	NR
D-dimer measurement		
Qualitative D-dimer positive	259 (43.3)	NR

	Modified/ simplified Wells (33) (n=3306)	**Original revised Geneva (34)** (n=965)	**Simplified revised Geneva (35)** (n=1049)
	1409 (42.6)	403 (41.8)	414 (39.5)
	53 (18)	61 (19)	56 (SD NR)
	NA	NA	241 (23.0)
	674 (20.4)	222 (23.0)	241 (23.0)
	190 (5.7)	NR	NR
	2032 (61.5)	NR	NR
	867 (26.2)	NR	NR
	610 (18.5)	232 (24.0)	NR
	480 (14.5)	166 (17.2)	NR
	176 (5.3)	43 (4.5)	NR
	476 (14.4)	89 (9.2)	NR
	NR	NR	NR
	NR	NR	NR
	NR	NR	NR
	NR	NR	NR
	NR	138 (14.3)	NR
	NR	NR	NR
	NR	51 (5.3)	NR
	NR	NR	NR
	NR	NR	NR
	NR	NR	NR
	NA§	NA	NA§

DVT=deep venous thrombosis; NA=not applicable; NR=not reported; bpm= beats per minute. *"Pain and swelling" is a composite variable, formed by variable "Oedema/swelling" present and/or variable "pain on deep vein (DV) palpation" present. †Missing values for "pain on DV palpation" (n=357 (59.7%)) were assigned as "pain on DV palpation" absent. ‡Missing values for "pain on DV palpation" (n=357 (59.7%)) were assigned as "pain on DV palpation" present. §Quantitative D-dimer test instead of qualitative point of care D-dimer test performed.

TABLE 2 HEAD TO HEAD COMPARISON OF C-STATISTICS OF FIVE DIAGNOSTIC PREDICTION MODELS (WITHOUT D-DIMER TESTING) TO RULE OUT PULMONARY EMBOLISM, VALIDATED IN PRIMARY CARE AMUSE-2 COHORT (N=598), USING DELONG METHOD (25)

	c-statistic			
Models compared	**Model 1**	**Model 2**	**Estimated difference (95% CI)**	**p-value**
Original Wells v simplified Wells	0.801	0.787	0.014 (-0.004 to 0.032)	0.132
Original Wells v modified Wells	0.801	0.798	0.003 (-0.001 to 0.007)	0.114
Original Wells v original revised Geneva	0.801	0.756	0.045 (-0.007 to 0.097)	0.091
Original Wells v simplified revised Geneva	0.801	0.748	0.053 (0.003 to 0.103)	0.038
Simplified Wells v modified Wells	0.787	0.798	-0.011 (-0.028 to 0.006)	0.203
Simplified Wells v original revised Geneva	0.787	0.756	0.031 (-0.017 to 0.079)	0.207
Simplified Wells v simplified revised Geneva	0.787	0.748	0.039 (-0.005 to 0.083)	0.084
Modified Wells v original revised Geneva	0.798	0.756	0.043 (-0.010 to 0.096)	0.113
Modified Wells v simplified revised Geneva	0.798	0.748	0.050 (0.000 to 0.100)	0.048
Original revised Geneva v simplified revised Geneva	0.756	0.748	0.008 (-0.010 to 0.026)	0.388

TABLE 3 DIAGNOSTIC ACCURACY MEASURES OF FIVE DIAGNOSTIC PREDICTION MODELS, COMBINED WITH POINT OF CARE D-DIMER TESTING, TO RULE OUT PULMONARY EMBOLISM, VALIDATED IN PRIMARY CARE AMUSE-2 COHORT (N=598), WITH (95% CIS)

Measure	**Original Wells ≤4**	**Modified Wells ≤2**	**Simplified Wells ≤1**	**Original revised Geneva ≤5***	**Simplified revised Geneva ≤2***
Sensitivity	95% (87% to 98%)	95% (87% to 98%)	96% (88% to 99%)	90% (81% to 96%)	88% (78% to 94%)
Specificity	51% (47% to 55%)	50% (46% to 55%)	49% (45% to 53%)	48% (44% to 53%)	53% (49% to 57%)
PPV	21% (17% to 26%)	21% (17% to 26%)	21% (17% to 25%)	20% (15% to 24%)	21% (16% to 26%)
NPV	99% (96% to 100%)	99% (96% to 100%)	99% (97% to 100%)	97% (95% to 99%)	97% (94% to 99%)
Efficiency	46% (41% to 50%)	45% (41% to 49%)	43% (39% to 48%)	44% (40% to 48%)	48% (44% to 52%)
Failure rate	1.5% (0.4% to 3.7%)	1.5% (0.4% to 3.8%)	1.2% (0.2% to 3.3%)	2.7% (1.1% to 5.4%)	3.1% (1.4% to 5.9%)

*Main analysis used composite variable "pain and swelling," in which missing values for "pain on deep vein (DV) palpation" were assigned as "pain on DV palpation" absent; results for analysis in which all missing values were assigned as "pain on DV palpation" present are identical and are not presented separately.
PPV= positive predictive value; NPV= negative predictive value.

TABLE 4 DIFFERENCES IN FAILURE RATES BETWEEN DIFFERENT DIAGNOSTIC PREDICTION MODELS IN AMUSE-2 COHORT (WITH 95% CIS BASED ON 200 BOOTSTRAP SAMPLES)

Comparison (model 1 vs. model 2)	**Difference between failure rates (model 1 vs. model 2)**
Original Wells vs. simplified Wells	0.32% (-0.08% to 1.04%)
Original Wells vs. modified Wells	-0.02% (-0.04% to 0.00%)
Original Wells vs. original revised Geneva	-1.21% (-2.49% to 0.14%)
Original Wells vs. simplified revised Geneva	-1.67% (-3.21% to -0.27%)
Simplified Wells vs. modified Wells	-0.33% (-1.06% to 0.06%)
Simplified Wells vs. original revised Geneva	-1.53% (-2.80% to -0.36%)
Simplified Wells vs. simplified revised Geneva	-1.98% (-3.33% to -0.74%)
Modified Wells vs. original revised Geneva	-1.20% (-2.47% to 0.18%)
Modified Wells vs. simplified revised Geneva	-1.65% (-3.21% to -0.25%)
Original revised Geneva vs. simplified revised Geneva	-0.45% (-1.23% to 0.21%)

EXTERNAL VALIDATION

In the validation cohort, venous thromboembolism was diagnosed in 73 of 598 patients (72 pulmonary embolism, 1 deep venous thrombosis) during the complete three month follow-up period (prevalence 12%). *Table 1* shows the baseline characteristics of the cohort, as well as the baseline characteristics observed in the development studies of the validated diagnostic prediction models. The main differences between the four development cohorts and the current validation cohort include the prevalence of the outcome, mean age, and percentage of male participants.

All models had moderate to good discriminative ability, with a c-statistic ranging from 0.75 (simplified revised Geneva score) to 0.80 (original and modified Wells rules) (largest difference P=0.038 (simplified revised Geneva v original Wells)) (see *Table 2* and Appendix *Figure B*).

For the originally suggested thresholds of the three Wells rules (original Wells, modified Wells, and simplified Wells, all with D-dimer testing if low predicted probability of pulmonary embolism), sensitivity was around 95%; it was slightly lower for the Geneva score models (88-90%). All five diagnostic models showed a specificity of approximately 50% (see *Table 3*). The simplified revised Geneva model was observed to be the most efficient rule (48%; 287 non-referred patients in cohort of 598 patients with suspected pulmonary embolism), but it was also associated with the highest failure rate (nine missed events in 287 non-referred patients; 3.1%). The largest difference in failure rates was observed between the simplified Wells and simplified revised Geneva score (-1.98%, 95% confidence interval -3.33% to -0.74%) (see *Table 4*). Interestingly, three of these nine patients were missed by all rules under study. These three patients are described in more detail in the box.

Overall, a one point lower cut-off (low risk original Wells 3 and D-dimer

TABLE 5 DIAGNOSTIC ACCURACY MEASURES (WITH 95% CIS) OF FIVE DIAGNOSTIC PREDICTION MODELS* COMBINED WITH POINT OF CARE D-DIMER TESTING, TO RULE OUT PULMONARY EMBOLISM, VALIDATED IN PRIMARY CARE AMUSE-2 COHORT (N=598)

	Original Wells		Modified Wells	
Measure	**≤5**	**≤3**	**≤3**	**≤1**
Sensitivity	89% (80% to 95%)	95% (87% to 98%)	89% (80% to 95%)	97% (91% to 99%)
Specificity	59% (54% to 63%)	50% (46% to 55%)	59% (54% to 63%)	32% (28% to 36%)
PPV	23% (18% to 28%)	21% (17% to 26%)	23% (18% to 28%)	17% (13% to 20%)
NPV	97% (95% to 99%)	99% (96% to 100%)	97% (95% to 99%)	99% (96% to 100%)
Efficiency	53% (49% to 57%)	45% (41% to 49%)	53% (49% to 57%)	28% (25% to 32%)
Failure rate	2.5% (1.1% to 4.9%)	1.5% (0.4% to 3.8%)	2.5% (1.1% to 4.9%)	1.2% (0.3% to 4.2%)

NPV=negative predictive value; PPV=positive predictive value.
*Chosen cut-off values for ruling out pulmonary embolism ("low PE probability") are 1 point higher or lower than cut-offs recommended for published models.

negative) affected the failure rate of each model little (failure rate 0.9-2.9%) but hampered its efficiency, especially for the simplified Wells rule (see *Table 5*). Conversely, an increased cut-off was more efficient, ranging from 49% to 54%, but more pulmonary embolism events were missed.

As can be appreciated from Appendix *Figure C*, the calibrations of the Wells rules and the revised Geneva score were good and comparable. Calibration of the simplified revised Geneva score was slightly worse, especially in the highest tenth of predicted risk.

DISCUSSION

Our systematic review identified 10 previously developed diagnostic prediction models to rule out pulmonary embolism. Of these, we evaluated five models for their transportability to primary care in an independent cohort of 598 patients. We found that all five models could rule out pulmonary embolism in about four in every 10 patients. However, we observed a difference in favour of the three Wells rules compared with the revised Geneva models in terms of safety. The proportion of cases of pulmonary embolism in those patients identified as being at low risk was substantially lower for the Wells rules.

Simplified Wells		Original revised Geneva		Simplified revised Geneva	
≤2	=0	≤6	≤4	≤3	≤1
85% (75% to 92%)	99% (93% to 100%)	89% (80% to 95%)	92% (83% to 97%)	86% (76% to 93%)	93% (85% to 98%)
59% (55% to 64%)	21% (18% to 25%)	54% (44% to 53%)	38% (34% to 42%)	59% (55% to 64%)	36% (32% to 41%)
23% (18% to 28%)	15% (12% to 18%)	21% (17% to 26%)	13% (10% to 16%)	23% (18% to 28%)	17% (13% to 21%)
97% (94% to 98%)	99% (95% to 100%)	97% (95% to 99%)	97% (94% to 99%)	97% (94% to 99%)	97% (94% to 99%)
54% (50% to 58%)	19% (16% to 22%)	49% (45% to 53%)	34% (30% to 38%)	54% (50% to 58%)	33% (29% to 37%)
3.4% (1.7% to 6.0%)	0.9% (0.0% to 4.9%)	2.8% (1.2% to 5.5%)	2.9% (1.1% to 6.3%)	3.1% (1.5% to 5.6%)	2.6% (0.8% to 5.9%)

STRENGTHS AND LIMITATIONS

Strengths of this study are that—to the best of our knowledge—this is the first study to validate multiple diagnostic strategies for ruling out pulmonary embolism in the primary care setting. Furthermore, the validated models were selected on the basis of an extensive literature search and critical appraisal of the models. For this, we used the guidance of the CHARMS checklist of the Cochrane Collaboration. (21) Moreover, we focused on the clinical performance measures of efficiency and failure rate; we believe that our results are relevant to daily practice, as these measures reflect the effect on clinical practice of using a diagnostic prediction model. Nevertheless, for full appreciation of our results, some limitations need to be considered.

Firstly, our study was conducted in the Netherlands, a country with a well-developed primary care structure in which general practitioners are the healthcare gatekeepers. Although in many west European countries, Canada, Australia, New Zealand, and parts of the United States, general practitioners play a rather comparable role in the healthcare system, our results may be less generalizable to healthcare settings where primary care medicine is less well developed.

Secondly, our case record form was primarily designed to score the items of the original Wells pulmonary embolism rule with the aim of validating this diagnostic model in a primary care setting. Accordingly, participating general practitioners

were not asked to score all items of other diagnostic prediction models at the moment of inclusion, although results for the scores of the other models could be calculated post hoc. However, in almost 60% of patients in our cohort data were lacking on pain on deep vein palpation, which is part of the Geneva models. In the main analysis, we assigned all missing values to "no pain." In a sensitivity analysis, we repeated all analyses with all missing values considered as "pain present." Results did not alter, indicating a limited influence of the frequently missing item in the whole prediction rule.

Thirdly, five more prediction models have been developed but were not validated in this study. Two models did not meet our a priori defined primary care criteria. Relying heavily on predictors not frequently assessed in primary care, such chest radiography results, the applicability of these models in primary care is limited. These predictors were not routinely collected in our validation cohort. Also, we did not have information on oxygen saturation and fever. Following guidelines on validation of prediction models, we wanted to refrain from simplification of the models by excluding some predictors and then refitting the new model on our data, as that would lead to the development of even more models. Simply omitting these missing predictors from the full models without further refitting would lead to structural underestimation of risks and poor clinical performance. Therefore, we refrained from validation of these five diagnostic prediction models in our primary care cohort.

Fourthly, debate is ongoing about the accepted failure rates in clinical practice. Although no true consensus exists, some people consider an upper 95% confidence interval boundary of a failure rate higher than 3% to be too high, although previous landmark studies in the field of diagnosing pulmonary embolism found failure rates with an upper 95% confidence interval boundary close to 4%. (20) Others focused on the point estimate of the failure rate instead, which should be under 2% given that even the most invasive diagnostic procedure (digital subtraction angiography) cannot diagnose all pulmonary embolism events and misses about 2% of cases. (39) We observed failure rates of the Wells rules of 1.2-1.5%, which is well below this 2% point estimate but with a 95% confidence interval that crosses 3%. Importantly, however, these boundaries for the 95% confidence interval are not necessarily the "true" values but reflect the remaining uncertainty around the point estimate.

Fifthly, for scoring the item "pulmonary embolism most likely", Wells originally suggested that information on electrocardiography, routine laboratory tests, and chest radiographs would be needed. These items are often not readily available in primary care, which may hamper the scoring of this subjective item in the Wells rule in primary care. Given that we observed good results for the Wells rule in primary care, it would be of interest to know how general practitioners actually interpret this subjective item. One hypothesis could be that this subjective item reflects the presence or absence of risk factors for venous thromboembolism, such as recent long haul flights or oestrogen use. However, when comparing the distribution of known risk factors for the subjective item scored or not, we

observed no clear differences (data not shown). Although this remains speculative, we believe that this subjective item "pulmonary embolism most likely" may reflect contextual knowledge from general practitioners on how symptoms are usually presented by their patients. Qualitative studies suggested that a consideration of "pulmonary embolism most likely" might come into the mind if a given patient presents symptoms differently ("out of the ordinary") compared with previous consultations. (1) Such contextual knowledge is often stronger in primary care, given the longstanding relationship that general practitioners often have with their patients.

Sixthly, despite the fact that most patients with the eventual diagnosis of pulmonary embolism were identified as having a high probability of pulmonary embolism by the validated diagnostic prediction models, three patients were missed by all of the rules. As shown in the box, these patients' characteristics were diverse, but two were young women who were taking oral contraception.

Finally, we used a qualitative point of care D-dimer test in this study. These tests are known to have a relatively lower sensitivity than laboratory based quantitative tests. (40, 41) Potentially, this resulted in a higher number of false negative results. However, specificity of point of care tests is higher, which favours the efficiency of the test. Moreover, the ease of performing a point of care test on the spot and having the results available within 15 minutes are convenient in a primary care setting. The trade-off between harms (increase in false negatives) and benefits (increased efficiency, quick point of care testing) should ideally be assessed in a cost effectiveness analysis.

RELATION TO DEVELOPMENT AND VALIDATION STUDIES

Multiple studies evaluated the diagnostic performance of the five diagnostic prediction models under study in a secondary (or tertiary) care setting. The failure rates observed in our validation study in primary care are largely in line with the previous studies regarding the performance of the Wells rules in combination with D-dimer testing, as can be appreciated from *Figure 3*: most secondary care studies yielded a failure rate of the diagnostic approach of combining the Wells rule with D-dimer testing around 1% (range 0.5-2.2%), whereas in our study we observed a failure rate close to 1.5% for the Wells rules combined with D-dimer testing.

Contrastingly, for the (simplified) revised Geneva scores, the failure rates observed in this primary care cohort are slightly higher than those in the previous studies (see *Figure 3*). Reasons for this are not entirely clear, but one can think of the differences in cohort characteristics that might have influenced the diagnostic performance. Importantly, the revised Geneva scores have been developed in a tertiary care setting, with a pre-selected and thus more severe disease presentation. Conversely, an unselected patient population and a lower prevalence of pulmonary embolism are typical features of a primary care setting, and such differences in case mix of patients can influence the diagnostic performance of the rule.

CLINICAL IMPLICATIONS & CONCLUSIONS

On the basis of these findings, we think that the revised Geneva scores are less suitable for use in this particular primary care setting, specifically when compared with the Wells rules. We would suggest that general practitioners use the (simplified version of the) Wells rule, combined with a (point of care) D-dimer test. Pulmonary embolism can be excluded in about four in every 10 patients with suspected pulmonary embolism, with an acceptably low failure rate of below 2%.

ACKNOWLEDGEMENTS

We thank AMUSE-2 project members R Oudega, H ten Cate, and M H Prins for their contribution to the design and initiation of the AMUSE 2 cohort. We thank Joris de Groot for his support with using the DeLong method and Peter Zuithoff and Karlijn Groenewegen for their statistical input.

BOX CHARACTERISTICS OF THREE CASES MISSED BY ALL FIVE DIAGNOSTIC PREDICTION MODELS COMBINED WITH D-DIMER TESTING WHEN VALIDATED IN AMUSE-2 COHORT, CONSISTING OF 598 PRIMARY CARE PATIENTS WITH SUSPECTED PULMONARY EMBOLISM

Case 1– 75 year old man
Venous thromboembolism in history; heart rate 59 bpm
Simplify D-dimer negative
Scores of models:

- Original Wells –1.5 (previous PE)
- Modified Wells –1.0 (previous PE)
- Simplified Wells –1.0 (previous PE)
- Original revised Geneva –4.0 (age ›65, previous PE)
- Simplified revised Geneva –2.0 (age ›65, previous PE)

Case 2– 30 year old woman
Use of oral contraceptives; heart rate 92 bpm
Simplify D-dimer negative
Scores of models:

- Original Wells –3.0 (PE most likely diagnosis)
- Modified Wells –2.0 (PE most likely diagnosis)
- Simplified Wells –1.0 (PE most likely diagnosis)
- Original revised Geneva –3.0 (heart rate 75-94 bpm)
- Simplified revised Geneva –1.0 (heart rate 75-94 bpm)

Case 3– 25 year old woman
Use of oral contraceptives; heart rate 80 bpm
Simplify D-dimer negative
Scores of models:

- Original Wells –0.0
- Modified Wells –0.0
- Simplified Wells –0.0
- Original revised Geneva –3.0 (heart rate 75-94 bpm)
- Simplified revised Geneva –1.0 (heart rate 75-94 bpm)

PE=pulmonary embolism; bpm=beats per minute.

REFERENCES

1. Barais M, Morio N, Cuzon Breton A, Barraine P, Calvez A, Stolper E, et al. "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(5):e98112.
2. Schiff G, Hasan O, Kim S, Abrams R, Cosby K, Lambert B, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169(20):1881-7.
3. Prasad V, Rho J, Cifu A. The diagnosis and treatment of pulmonary embolism: a metaphor for medicine in the evidence-based medicine era. Arch Intern Med. 2012;172(12):955-8.
4. Wiener R, Schwartz L, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. Arch Intern Med. 2011;171(9):831-7.
5. Lencioni R, Fattori R, Morana G, Stacul F. Contrast-induced nephropathy in patients undergoing computed tomography (CONNECT)- a clinical problem in daily practice? A multicenter observational study. Acta Radiology. 2010;51(7):741-50.
6. Le Gal G, Bounameaux H. Diagnosing pulmonary embolism: running after the decreasing prevalence of cases among suspected patients. J Thromb Haemost. 2004;2(8):1244-6.
7. Lucassen W, Geersing G, Erkens PMG, Reitsma JB, Moons KGM, Büller H, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. Ann Intern Med. 2011;155(7):448-60.
8. Altman D, Vergouwe Y, Royston P, Moons K. Prognosis and prognostic research: validating a prognostic model. Bmj. 2009;338:b605.
9. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
10. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
11. Hendriksen JMT, Geersing GJ, Moons KGM, de Groot JAH. Diagnostic and prognostic prediction models. J Thromb Haemost. 2013;11:129-41.
12. Altman D, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19:453 - 73.
13. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-9.
14. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-24.
15. Knottnerus JA. Between iatrotropic stimulus and interiatric referral:: the domain of primary care research. Journal of Clinical Epidemiology. 2002;55(12):1201-6.
16. Oudega R, Hoes AW, Moons KGM. The Wells rule does not adequately rule out deep venous thrombosis in primary care patients. Ann Intern Med. 2005;143(2):100-7.
17. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085-94.
18. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68(3):279-89.

19. Geersing GJ, Erkens PMG, Lucassen W, Buller H, Ten Cate H, Hoes A, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
20. Wells P, Andersen D, Rodger M, Ginsberg JS, Kearon C, Gent M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. Thromb Haemost. 2000;83(3):416-20.
21. Moons KGM, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10).
22. Riley RD, Ridley G, Williams K, Altman DG, Hayden J, de Vet HCW. Prognosis research: toward evidence-based results and a Cochrane methods group. J Clin Epidemiol. 2007;60(8):863-5.
23. Collins GS, Moons KGM. Comparing risk prediction models. BMJ. 2012;344:e3186.
24. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. BMJ. 2009;339:b4184.
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
26. Douma R, le Gal G, Sohne M, Righini M, Kamphuisen PW, Perrier A, et al. Potential of an age adjusted D-dimer cut-off value to improve the exclusion of pulmonary embolism in older patients: a retrospective analysis of three large cohorts. BMJ. 2010;340.
27. Janssen KJ, Vergouwe Y, Donders AR, Harrell FE, Jr., Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. Clin Chem. 2009;55(5):994-1001.
28. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med. 1991;10(4):585-98.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:10.1136/bmj.g7594.
30. Moons KGM, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73.
31. Miniati M, Monti S, Bottai M. A structured clinical model for predicting the probability of pulmonary embolism. Am J Med. 2003;114(3):173-9.
32. 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(1):92-7.
33. Gibson NS, Sohne M, Kruip MJHA, Tick LW, Gerdes VE, Bossuyt PM, et al. Further validation and simplification of the Wells clinical decision rule in pulmonary embolism. Thromb Haemost. 2008;99:229-34.
34. Le Gal G, Righini M, Roy P-M, Sanchez O, Aujesky D, Bounameaux H, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. Ann Intern Med. 2006;144(3):165-71.

35. Klok FA, Mos IM, Nijkeuter M, et al. Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism. Arch Intern Med. 2008;168(19):2131-6.
36. Miniati M, Bottai M, Monti S, Salvadori M, Serasini L, Passera M. Simple and accurate prediction of the clinical probability of pulmonary embolism. Am J Respir Crit Care Med. 2008;178(3):290-4.
37. Kline JA, Mitchell AM, Kabrhel C, Richman PB, Courtney DM. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. J Thromb Haemost. 2004;2(8):1247-55.
38. 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(2):144-52.
39. Perrier A, Roy PM, Sanchez O, Le Gal G, Meyer G, Gourdier AL, et al. Multidetector-row computed tomography in suspected pulmonary embolism. N Engl J Med. 2005;352:1760-8.
40. Geersing GJ, Janssen KJM, Oudega R, Bax L, Hoes AW, Reitsma JB, et al. Excluding venous thromboembolism using point of care D-dimer tests in outpatients: a diagnostic meta-analysis. BMJ. 2009;339(b2880).
41. Stein PD, Hull RD, Patel KC, Olson RE, Ghali WA, Brant R, et al. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. Ann Intern Med. 2004;140(8):589-602.
42. van Belle B, Buller H, Huisman M, Kaasjager K, Kamphuisen P. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. JAMA. 2006;295(2):172-9.
43. Perrier AMD, Desmarais S, Miron MJ, de Moerloose P, Lepage R, Slosman D, et al. Non-invasive diagnosis of venous thromboembolism in outpatients. Lancet. 1999;353(9148):190-5.
44. Perrier A, Desmarais S, Goehring C, De Moerloose P, Morabia A, Unger PF, et al. D-dimer testing for suspected pulmonary embolism in outpatients. Am J Respir Crit Care Med. 1997;156(2):492-6.

APPENDIX BOX A DESIGN OF THE SYSTEMATIC REVIEW ON DIAGNOSTIC PREDICTION MODELS FOR DIAGNOSIS OF SUSPECTED PULMONARY EMBOLISM, BASED ON THE CHARMS CHECKLIST (21)

Item	
Scope of the review	Review of all existing prediction models for diagnosis of pulmonary embolism (PE), and subsequent external validation of these models in an independent cohort of 598 primary care patients suspected of PE. The models are intended to inform physicians on referral to secondary care, or withholding from invasive diagnostic testing based on the model's estimated probability of having PE. In case of low probability: no referral to secondary care or no invasive diagnostic testing.
Type of primary studies	Diagnostic prediction model development studies with or without external validation in independent data.
Target population	Primary care patients in whom the diagnosis pulmonary embolism is considered: - Unexplained acute dyspnoea, and/ or - Unexplained cough, and/ or - Pain on inspiration
Outcome to be predicted	Pulmonary embolism present or absent as determined by an established reference standard, such as spiral CT scanning, pulmonary angiography, ventilation-perfusion scanning, clinical follow-up or a combination of these.

APPENDIX TABLE A DETAILS OF THE RETRIEVED DERIVATION STUDIES BASED ON THE CHARMS CHECKLIST (21)

	original Wells rule (20)	**simplified & modified Wells rule (33)**	**Geneva score (32)**	**revised Geneva score (34)**
Objective	Derivation diagnostic prediction model	Simplification of Wells rule	Derivation diagnostic prediction model	Derivation diagnostic prediction model
Source of data	Prospective cohort	Prospective cohort	2 prospective cohorts	Prospective cohort
Participants	- consecutive in- and outpatients with suspected PE - inclusion period NR - hospital setting: NR	- see (42) - consecutive in- and outpatients with suspected PE - inclusion Nov 2002- Aug 2004 - 12 Dutch hospitals (5 academic, 7 general urban)	- see (43, 44) - consecutive patients presenting at ER with suspected PE - inclusion Oct 1 1992- Oct 31 1997. - 1 academic hospital	- consecutive patients presenting at ER with suspected PE - inclusion Oct 2000 – Jun 2002 - 3 hospitals in Switzerland and France
Outcomes to be predicted	- PE present/ absent - imaging in all patients - 3 months follow-up - outcome assessment blinded for other information	- PE present/ absent - DPM & D-dimer for exclusion, CT in high risk and/or positive D-dimer - 3 months follow-up - final diagnosis: trained radiologist - blinding: NR	- PE present/ absent - sequence of non-invasive instruments (including clinical assessment, lung scan, ELISA D-dimer, lower limb CUS) - angiogram if inconclusive workup - blinding: NR	- PE present/ absent - DPM & negative D-dimer (VIDAS) for exclusion, imaging if high risk - 3 months follow-up - blinding: NR
Candidate predictors	- 40 candidate predictors - patients history, physical examination, additional testing. - dichotomization of continuous variable - predictor measurement at patient presentation, blinded for final diagnosis (outcome)	- predictors of original Wells rule - predictor assessment blinded for outcome: NR	- 30 candidate predictors - patient characteristics, risk factors VTE, symptoms, signs, X-thorax, blood gas analysis - blinded for final diagnosis - dichotomization and categorization of continuous variables	- 26 candidate predictors - patient history and physical examination predictors, no subjective items - predictor categories: age, heart rate - predictor measurement at patient presentation, blinded for final diagnosis (outcome)

simplified revised Geneva score (35)	Charlotte rule (38)	PERC (37)	Pisa rule (31)	revised Pisa rule (36)
Simplification of revised Geneva score	Derivation diagnostic prediction model	Derivation diagnostic prediction model	Derivation diag pred model	Simplification of Pisa Rule (no X-thorax)
2 prospective cohorts	Prospective cohort	Prospective cohort	Prospective cohort	Prospective cohort
Study A (39): consecutive patients presenting at ER with suspected PE. - inclusion Aug 2002- Nov 2003 - 3 hospitals (France& Swiss). Study B (42): conse-cutive in-outpatients with suspected PE - inclusion Nov 2002- Aug 2004 - 1 Dutch 3rd centre	- patients presenting at ER, (highly) suspected PE. PE work-up already planned - inclusion 1996- 2000. - 7 ERs in United States	- patients presenting at ER, evident clinical suspicion of PE - inclusion period NR - 10 ERs in United States	- consecutive patients referred to hospital with suspected PE: - inclusion Nov 1 1991- Dec 31 1999 - 70% referred from medical/ surgical departments & ER; 30% of 4 peripheral hospitals	- consecutive patients referred to hospital with suspected PE: - inclusion Nov 1 1991- Dec 31 1999 - 70% referred from medical/ surgical departments& ER; 30% of 4 peripheral hospitals
- Study A: see (39) - Study B: see (42)	- PE present/ absent - imaging in all patients (angiography, VQ scan, CT scanning) , or autopsy - 6 months follow-up - interpretation of imaging by assessors blinded for other information	- PE present/ absent - in 650 patients PE ruled out by structured protocol with D-dimer & alveolar dead space measurement - 90 days follow-up - remaining patients diagnosis by imaging - blinding: NR	- PE present/ absent - imaging (angio-graphy, perfusion scan) or autopsy in all patients - 6 months follow-up - blinding: NR	- PE present/ absent - imaging (angiography, perfusion scan) or autopsy in all patients - 6 months follow-up - outcome assessment blinded to clinical information
- predictors of revised Geneva score - predictor assessment blinded for outcome	- 26 candidate predictors - patients' history, physical examination, additional testing - continuous variables dichotomized if included in model - predictor assessment blinded for outcome: NR	- 21 candidate predictors - patient characteristics - assessment at patient presentation, before diagnosis; 4 items retrospectively assessed - dichotomization of significant continuous variables	- 34 candidate predictors - clinical history, physical examination, ECG/ X-thorax, PaO2, PaCO2 - standardized form before further objective testing - split of continuous variables into tertiles	- 16 candidate predictors - see Pisa rule, without blood gas or X-thorax - standardized form before further objective testing - split of continuous variable age into quartiles

APPENDIX TABLE A EXTENDED

	original Wells rule (20)	simplified & modified Wells rule (33)	Geneva score (32)	revised Geneva score (34)
Sample size	- 1,260 suspected patients included. - PE present in 222 patients (17.6%) - EPV <10 (222/40)	- 3,306 suspected patients included. - PE present in 674 patients (20%) - EPV n.a	- 1,090 suspected patients included - PE present in 296 patients (27%) - EPV ≈10	- 1,280 suspected patients screened, 965 patients included - PE present in 222 patients (23.0%) - EPV <10 (222/26)
Missing data	- no D-dimer result in 49 patients - no info on other missing data	- D-dimer missing in 2% of all patients with low probability (≤ 4) - DPM score in 3,298 patients (99.8%)	- exclusion of 3 patients with missing data (details NR) - 104 patients with missing data - DPM developed in 986 patients	- number of participants with missing data for each predictor reported - predictor exclusion if >2% missing data
Model development	- multivariable logistic regression - predictor selection: univariable regression <0.15 - stepwise regression with p<0.05 during multivariable modelling - no shrinkage of predictor weights	- evaluation of simplified & modified Wells rules by assigning 1 or 2 point(s) per item if present	- multivariable logistic regression - predictor selection: univariable regression <0.05 - full model approach with p<0.05 multivariable modelling - cross-validation procedure to examine degree overfitting: substantial overfitting was ruled out	- multivariable logistic regression - predictor selection: univariable regression <0.05 - removal of non-statistically significant variables from model - no shrinkage of predictor weights
Model performance	- discrimination and calibration: NR - comparison of predictive values using different cut-offs	- discrimination: AUC ROC curve - calibration: NR - a priori cut-offs used	- discrimination: comparison of AUCs for naïve & cross-validated scores - calibration: NR - comparison with empirical assessment by ED physician	- discrimination: AUC ROC curve - calibration: Hosmer-Lemeshow P=0.55 - predicted-observed table

simplified revised Geneva score (35)	Charlotte rule (38)	PERC (37)	Pisa rule (31)	revised Pisa rule (36)
- 1,049 suspected patients included. - PE present in 241 patients (23.0%) - EPV n.a.	- 934 suspected patients included. - PE present in 181 patients (19.4%). - EPV <10	- 3,148 suspected patients included - PE present in 348 patients (11.0%) - EPV >10	- 1,100 suspected patients included. - PE present in 440 patients (40%) - EPV >10	- 1,100 suspected patients included - PE present in 440 patients (40%) - EPV >10
- study A& B: minimal loss-to–follow-up - missing D-dimer data: 2.4% in low risk, 9.5% in intermediate risk, 4.3% in PE-unlikely group	- no clear description of variables missing per participants - exclusion if >5% missing (arterial blood gas analysis)	- missing data: none	- missing data: NR	- missing data: NR
- evaluation of revised Geneva score by assigning 1 point per item if present	- multivariable logistic regression - predictor selection: exclusion if underrepresentation, poor inter-observer reliability, missing data, or if ≥2 predictors can be collapsed into 1 - if significant in multivariable model (P<0.05) used in decision tree - shrinkage: NR	- multivariable logistic regression - all candidate variables in model - variable selection via modified backward stepwise process: exclusion of categorical & dichotomized variables with lower 95% CI bound for Cohen's K <0.40 - shrinkage: NR	- multivariable logistic regression - predictor selection: backward selection P>0.20. Remaining variables kept in model if individually statistically significant. Forward selection of removed variables, kept in model if statistically significant or deemed to be confounder - pairwise interactions tested	- multivariable logistic regression - predictor selection: backward selection. Remaining variables kept in model if individually statistical significant. If change coefficients >10% after removal, reintroduction variable - age & sex kept in model regardless of statistical significance
- discrimination AUC ROC curve. - calibration: NR	- discrimination: NR - calibration: goodness-of-fit by Hosmer-Lemeshow test	- discrimination: NR - calibration: model fit: likelihood ratio chi-squared, Hosmer-Lemeshow - diagnostic performance in 2 populations - no reclassification or NRI	- discrimination: AUC final model, with 95% CIs by 1000 bootstrap samples - calibration: NR	- discrimination: AUC final model, with 95% CIs by 1000 bootstrap samples - calibration: NR

APPENDIX TABLE A EXTENDED

	original Wells rule (20)	simplified & modified Wells rule (33)	Geneva score (32)	revised Geneva score (34)
Model evaluation	- internal validation: random split-sample; 80% derivation set, 20% validation set - external validation: NR - no further adjustment or update	n.a.	- internal validation: cross-validation - external validation: NR - no further adjustment or update	- internal validation: random split-sample 90% derivation, 10% validation set; procedure 10x repeated - external validation: independent cohort (temporal) - no further adjustment or update
Results	- final model with original regression coefficients, odds ratios & rounded predictor weights - intercept NR	- original regression coefficients and odds ratios reported - intercept NR - no comparison of distribution predictors in derivation & validation set	- final model with regression coefficients - intercept: NR - model with rounded predictor weights reported	- final model with original regression coefficients and rounded predictor weights - intercept NR. - comparison of distribution different predicted risk groups for derivation & validation set

NR= not reported; n.a.= not applicable; PE= pulmonary embolism; DPM= diagnostic prediction model; EPV = events per variable; AUC= area under the curve; ROC curve= receiver operating characteristics curve; 95% CI= 95% confidence interval

	simplified revised Geneva score (35)	Charlotte rule (38)	PERC (37)	Pisa rule (31)	revised Pisa rule (36)
	n.a.	- internal validation by generating 95% CIs for the odds ratios using bootstrap - external validation: NR	- internal validation: NR - external validation in 2 populations - no further adjustment or update	- internal validation: cross-validation 90% derivation set; 10% validation set, procedure 10x - external validation NR	- internal validation: assessment overall accuracy (AUC ROC) estimated in 1000 bootstrap samples - external validation: independent sample of 400 patients ('03-'05)
	- new regression coefficients reported - intercept: NR - no comparison of distribution predictors in derivation & validation set	- decision tree based on significant factors of multivariable model	- final model with intercept & regression coefficients - block rule presentation	- final model with intercept & regression coefficients - graph to estimate probability of PE - comparison probability estimates and actual PE prevalence	- final model with intercept & regression coefficients - comparison probability estimates and actual PE prevalence

APPENDIX- TABLE B TRIPOD CHECKLIST

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	✓
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	✓
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	✓
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	✓
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	✓
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	✓
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	✓
	5b	D;V	Describe eligibility criteria for participants.	✓
	5c	D;V	Give details of treatments received, if relevant.	n.a.
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	✓
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	✓
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	✓
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	✓
Sample size	8	D;V	Explain how the study size was arrived at.	✓
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	✓
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	n.a.
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	n.a.
	10c	V	For validation, describe how the predictions were calculated.	✓
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	✓
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	n.a.

APPENDIX- TABLE B EXTENDED

Section/Topic			Checklist Item	Page
Results				
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	✓
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	✓
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	✓
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	✓
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	✓
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	n.a
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	n.a.
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	n.a.
	15b	D	Explain how to the use the prediction model.	n.a.
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	✓
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	n.a.
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	✓
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	✓
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	✓
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	✓
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	✓
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	✓

APPENDIX TABLE C OVERVIEW OF ALL DIAGNOSTIC PREDICTION MODELS TO RULE OUT PE IDENTIFIED BY THE SYSTEMATIC LITERATURE SEARCH.

Wells rule (20, 33)

		Original	Modified	Simplified
	RC	Points		
Clinical signs & symptoms DVT	1.8	3.0	2	1
Alternative diagnosis less likely	1.5	3.0	2	1
Heart rate › 100 bpm	1.1	1.5	1	1
Immobilization	0.92	1.5	1	1
Previous VTE	0.87	1.5	1	1
Haemoptysis	0.87	1.0	1	1
Malignancy	0.81	1.0	1	1
Cut-off for PE unlikely		*≤4*	*≤2*	*≤1*

RC= regression coefficient; DVT= deep venous thrombosis; bpm = beats per minute; VTE= venous thromboembolic event.

Geneva score (32)

		Original
	regression coefficients	Points
Age 60-79 years	0.6	1
≥80	1.0	2
Previous VTE	1.1	2
Surgery	1.5	3
Pulse rate ›100 bpm	0.5	1
PaCO2, kPa		
‹4.8	1.1	2
4.8-5.19	0.6	1
PaO2, kPa		
‹6.5	2.0	4
6.5- 7.99	1.4	3
8- 9.49	1.0	2
9.5- 10.99	0.6	1
Chest X-ray		
Platelike atelectasis	0.7	1
Elevation of hemidiaphragm	0.5	1
Cut-off clinical probability	*Low*	*‹5*
	Intermediate	*5-8*
	High	*›8*

APPENDIX TABLE C EXTENDED

revised Geneva score (34, 35)

		Original	Simplified
	RC	Points	
Age >65 years	0.39	1	1
Previous VTE	1.05	3	1
Surgery/ bone fracture	0.78	2	1
Malignancy	0.45	2	1
Unilateral lower limb pain	0.97	3	1
Haemoptysis	0.74	2	1
Heart rate			
75-94 bpm	1.20	3	1
≥ 95 bpm	0.67	5	1
Pain on deep venous palpation and unilateral oedema	1.34	4	1
Cut-off for PE unlikely		*≤5*	*≤2*

RC= regression coefficient; DVT= deep venous thrombosis; bpm = beats per minute; VTE= venous thromboembolic event.

PERC (37)

For a negative result, the clinician must answer "no" to the following 8 questions:

- Is the patient aged >49 y?
- Is the pulse >99 beats/min?
- Is the pulse oximetry reading <95% while the patient breathes room air?
- Is there a history of haemoptysis?
- Is the patient receiving exogenous oestrogen?
- Does the patient have a previous diagnosis of venous thromboembolism?
- Has the patient had recent surgery or trauma that required endotracheal intubation or hospitalization in the previous 4 weeks?
- Does the patient have unilateral leg swelling (on the basis of visual observation of asymmetry of the calves)?

Charlotte rule (38)

This rule classifies patients as either safe (eligible for D-dimer testing) or unsafe.

- If the patient is aged ≤50 y and his or her heart rate is less than or equal to their systolic blood pressure (shock index ≤1.0), the patient is safe.
- If the patient is aged >50 y or has a shock index >1.0, the clinician should ask 4 sequential questions:

- Does the patient have unexplained hypoxemia?
- Does the patient have unilateral leg swelling?
- Has the patient had surgery that required general anaesthesia in the past 4 weeks?
- Does the patient have haemoptysis?

If the answer to all 4 questions is "no," then the patient is safe.

APPENDIX TABLE C EXTENDED

Pisa rule (31)

		Regression coefficients
Male sex		0.81
Age		
63-72		0.59
≥73		0.92
Pre-existing		
Cardiovascular disease		-0.56
Pulmonary		-0.97
Thrombophlebitis		0.69
Symptoms		
Dyspnoea		1.29
Chest pain		0.64
Haemoptysis		0.89
Temp >38°C		-1.17
ECG sings of acute right ventricular overload		1.53
Findings X-thorax		
Oligemia		3.86
Amputation of hilar artery		3.92
Consolidation (infarction)		3.55
Consolidation (non-infarction)		-1.23
Pulmonary Oedema		-2.83
Constant		-3.26
Clinical probability:	*Low*	*0-10%*
	Intermediate	*11-50%*
	Moderately high	*51-90%*
	High	*>90%*

APPENDIX TABLE C EXTENDED

revised Pisa rule (36)

		Regression coefficients
Age		
57-67		0.80
68-74		0.87
≥75		1.14
Male sex		0.60
Immobilization		0.42
Pre-existing		
Cardiovascular disease		-0.51
Pulmonary		-0.89
Thrombophlebitis		0.64
Symptoms		
Dyspnoea		2.00
Orthopnoea		-1.51
Chest pain		1.01
Fainting/ syncope		0.66
Haemoptysis		0.93
Leg Swelling		0.80
Temp >38°C		-1.47
Wheezes		-1.20
Crackles		-0.61
Acute cor pulmonale on ECG		1.96
Constant		-3.43
Clinical probability:	*Low*	*0-10%*
	Intermediate	*11-50%*
	Moderately high	*51-80%*
	High	*81-100%*

APPENDIX FIGURE A SYSTEMATIC LITERATURE SEARCH IN PUBMED & EMBASE DATABASES

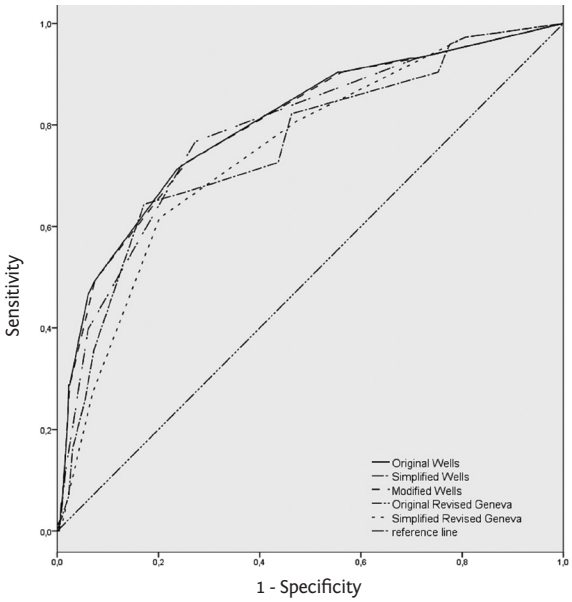
PubMed 13-10-2014

("Pulmonary Embolism"[Mesh:noexp] OR (lung embolism[tiab] OR lung embolisms[tiab]) OR (pulmonary embolism[tiab] OR pulmonary embolism/clinical[tiab] OR pulmonary embolism/infarction[tiab] OR pulmonary embolisms[tiab]) OR (pulmonary thromboembolism[tiab] OR pulmonary thromboembolisms[tiab])) AND ("Epidemiologic Research Design"[Mesh] OR "Epidemiologic Studies"[Mesh] OR "Probability"[Mesh] OR probability[tiab] OR probabilities[tiab] OR "Outcome Assessment (Health Care)"[Mesh] OR ((validitat[tiab] OR validitate[tiab] OR validitates[tiab] OR validitation[tiab] OR validite[tiab] OR validited[tiab] OR validities[tiab] OR validiting[tiab] OR validition[tiab] OR validitiy[tiab] OR validity[tiab] OR validity/accuracy[tiab] OR validity/applicability[tiab] OR validity/biomarker[tiab] OR validity/clinical[tiab] OR validity/credibility[tiab] OR validity/diagnostic[tiab] OR validity/effective[tiab] OR validity/expresses[tiab] OR validity/feasibility[tiab] OR validity/generalizability[tiab] OR validity/invalidity[tiab] OR validity/no[tiab] OR validity/predictive[tiab] OR validity/quality[tiab] OR validity/relationship[tiab] OR validity/reliability[tiab] OR validity/representativity[tiab] OR validity/reproducibility[tiab] OR validity/scientific[tiab] OR validity/screening[tiab] OR validity/test[tiab] OR validity/utility[tiab] OR validity/validation[tiab] OR validity/viability[tiab] OR validity'[tiab] OR validity's[tiab] OR validityof[tiab]) AND and reliability[tiab]) OR (clinical prediction rule[tiab] OR clinical prediction rules[tiab]) OR "Sensitivity and Specificity"[Mesh] OR (roc curve[tiab] OR roc curve/area[tiab] OR roc curve/sensitivity[tiab] OR roc curves[tiab]) OR (pretest[All Fields] AND clinical[All Fields] AND model[All Fields])) AND (("2010/01/01"[PDAT] : "2014/10/01"[PDAT]) AND "humans"[MeSH Terms] AND (Dutch[lang] OR English[lang] OR German[lang] OR French[lang] OR Italian[lang] OR Spanish[lang]) NOT ("child"[MeSH Terms:noexp] OR "infant"[MeSH Terms:noexp] OR "infant"[MeSH Terms] OR "child, preschool"[MeSH Terms] OR "infant, newborn"[MeSH Terms])) NOT (Editorial[ptyp] OR Letter[ptyp] OR Meta-Analysis[ptyp] OR Practice Guideline[ptyp] OR Review[ptyp])

Embase 13-10-2014

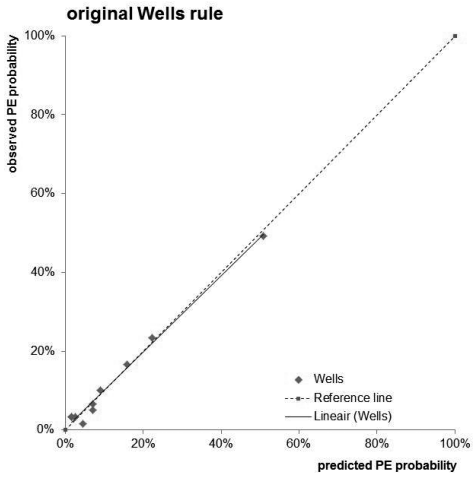
#18 #17 AND [abstracts]/lim
#17 #16 AND [embase]/lim AND [2010-2014]/py
#16 #15 AND #5
#15 #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15
#14 (pretest NEXT/1 clinical):ab,ti AND model*:ab,ti AND [embase]/lim
#13 (roc NEXT/1 curve*):ab,ti AND [embase]/lim
#12 'roc curve' AND [embase]/lim
#11 'sensitivity and specificity' AND [embase]/lim
#10 clinical:ab,ti AND prediction:ab,ti AND rule*:ab,ti AND [embase]/lim
#9 valid*:ab,ti AND reliab*:ab,ti AND [embase]/lim
#8 'outcome assessment' AND [embase]/lim
#7 probabilit*:ab,ti AND [embase]/lim
#6 'probability'/exp AND [embase]/lim
#5 #1 OR #2 OR #3 OR #4
#4 (pulmonary NEXT/1 thromboembolism*):ab,ti AND [embase]/lim
#3 (lung NEXT/1 embolism*):ab,ti
#2 (pulmonary NEXT/1 embolism*):ab,ti AND [embase]/lim
#1 'lung embolism' AND [embase]/lim

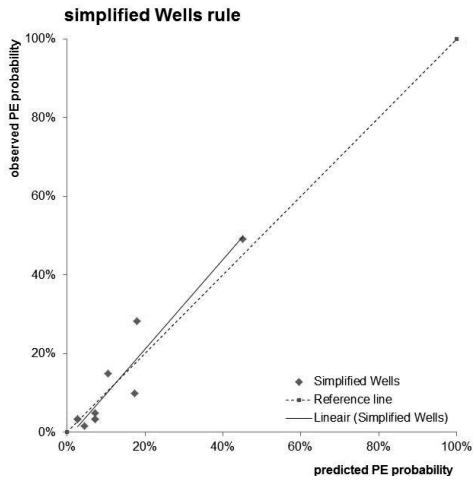
APPENDIX FIGURE B RECEIVER OPERATING CHARACTERISTICS (ROC) CURVES WITH ESTIMATED C-STATISTICS (95% CONFIDENCE INTERVALS) OF THE FIVE PREDICTION MODELS (WITHOUT D-DIMER TESTING) IN THE AMUSE-2 COHORT



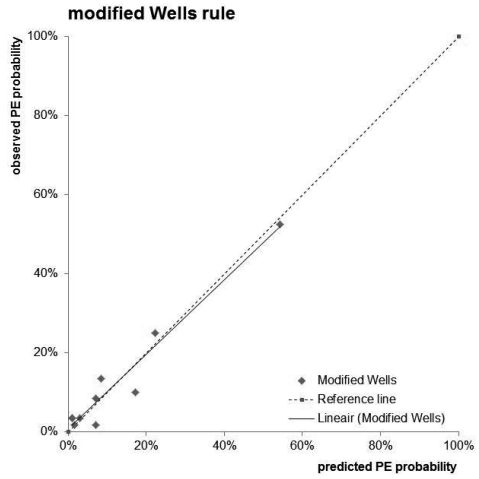
	c-statistic
Original Wells	0.80 (0.74 to 0.86)
Simplified Wells	0.79 (0.73 to 0.85)
Modified Wells	0.80 (0.74 to 0.86)
Original revised Geneva	0.76 (0.69 to 0.82)
Simplified revised Geneva	0.75 (0.69 to 0.81)

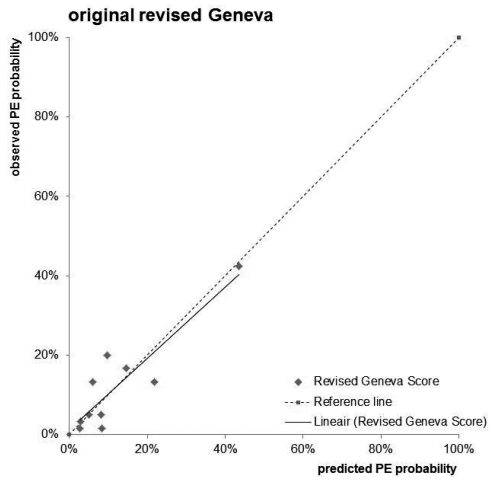
APPENDIX- FIGURE C CALIBRATION PLOTS OF THE ORIGINAL, MODIFIED AND SIMPLIFIED WELLS RULES AND THE ORIGINAL AND SIMPLIFIED REVISED GENEVA SCORES IN THE AMUSE-2 COHORT CONSISTING OF 598 PATIENTS SUSPECTED OF PULMONARY EMBOLISM IN PRIMARY CARE, BASED ON THE PREDICTED PROBABILITY OF PULMONARY EMBOLISM PRESENT BY CALCULATING THE LINEAR PREDICTOR FOR EACH OF THE MODELS.

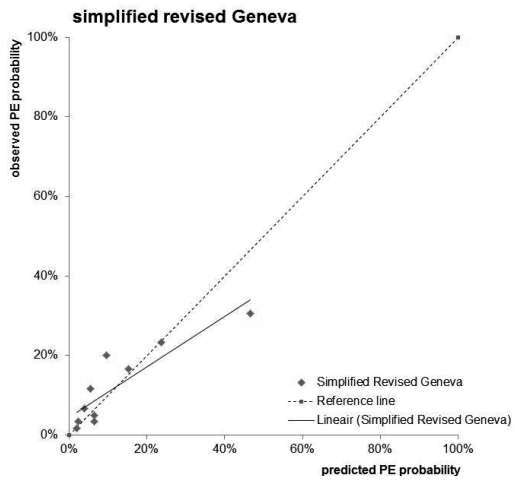


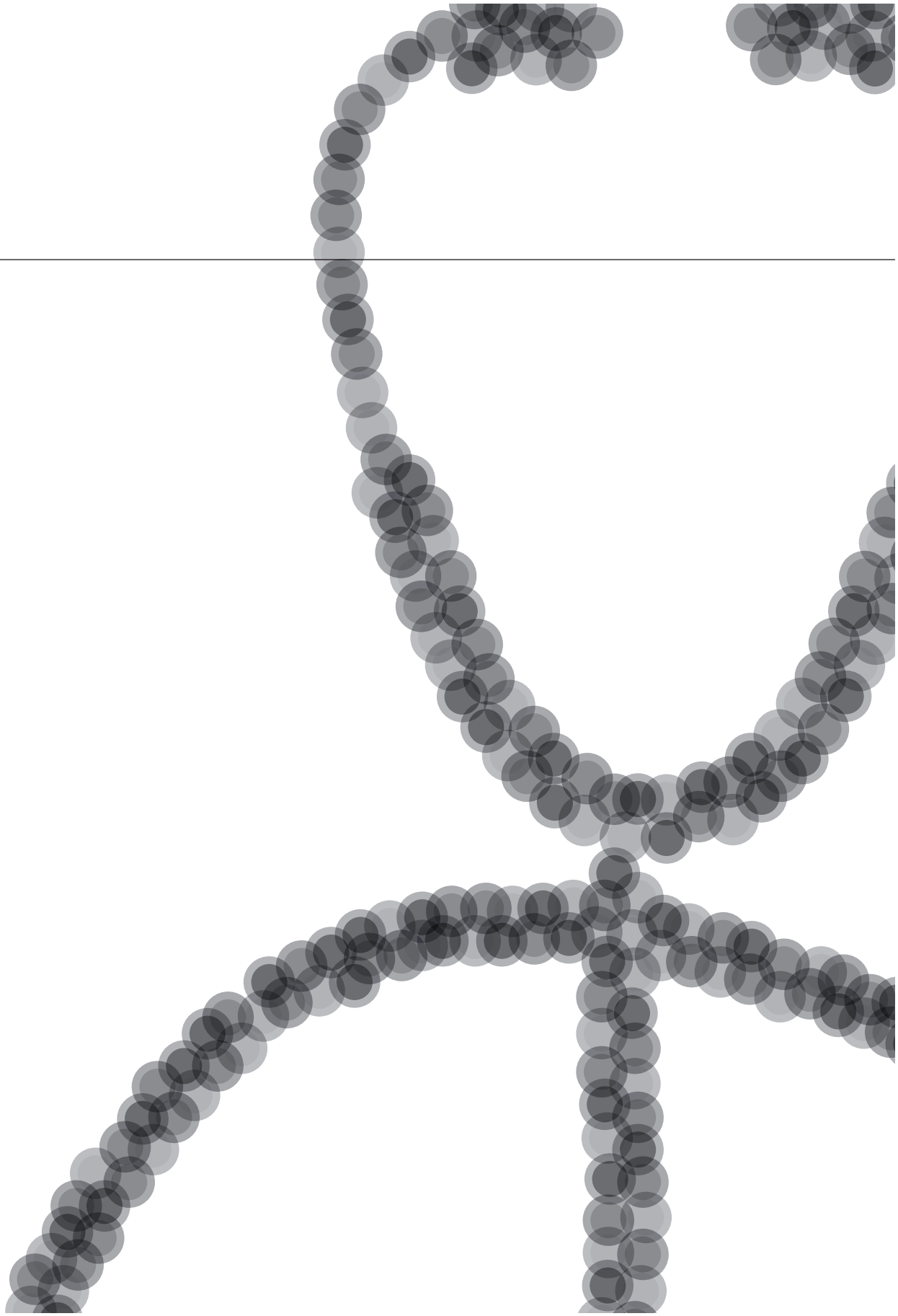


PE= pulmonary embolism


modified Wells rule
observed PE probability
100%
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
predicted PE probability
Modified Wells
Reference line
Lineair (Modified Wells)


original revised Geneva
observed PE probability
100%
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
predicted PE probability
Revised Geneva Score
Reference line
Lineair (Revised Geneva Score)

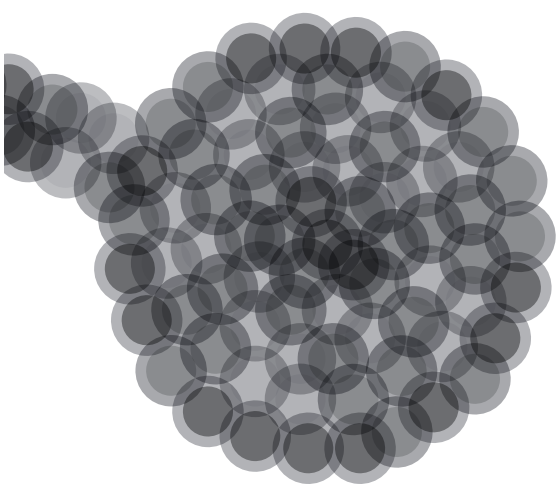

simplified revised Geneva
observed PE probability
100%
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
predicted PE probability
Simplified Revised Geneva
Reference line
Lineair (Simplified Revised Geneva)



CHAPTER 4

RULING OUT PULMONARY EMBOLISM IN PRIMARY CARE: COMPARISON OF THE DIAGNOSTIC PERFORMANCE OF "GESTALT" AND THE WELLS RULE

Janneke MT Hendriksen, Wim AM Lucassen, Petra MG Erkens,
Henri EJH Stoffers, Henk CPM van Weert, Harry R Büller,
Arno W Hoes, Karel GM Moons, Geert-Jan Geersing



Annals of Family Medicine;
provisionally accepted pending revisions

ABSTRACT

Purpose: Diagnostic prediction models, like the Wells-rule, can be used to safely rule-out pulmonary embolism (PE) in suspected patients. However, a physician's own probability estimate ("gestalt") is commonly used instead, despite its uncertain diagnostic value. We evaluated the diagnostic performance of both approaches in primary care.

Methods: Family doctors (FDs) estimated the probability of PE presence on a 0-100% scale ("gestalt") and calculated the Wells-rule in 598 patients suspected of PE. All patients were referred to secondary care for reference testing. We compared the discriminative ability (c-statistic) of both approaches. Next, we stratified patients into PE risk categories. For "gestalt" an estimated probability ‹20% plus a negative point-of-care D-dimer (Clearview Simplify) indicated low-risk. Similarly, for the Wells-rule we used a score of 4 plus a negative D-dimer. We compared sensitivity, specificity, efficiency (% low-risk patients in total cohort) and failure rate (% PE patients within low-risk category).

Results: Venous thromboembolism was confirmed in 73 patients (prevalence 12%). The c-statistic was 0.77 (95%CI 0.70-0.83) for "gestalt" and 0.80 (95%CI 0.75-0.86) for the Wells-rule. Using our pre-defined risk stratification, "gestalt" missed 2 out of 152 low-risk patients (failure rate 1.3%, 95%CI 0.2-4.7%) with an efficiency of 25% (95%CI 22-29%); the Wells-rule missed 4 out of 272 low-risk patients (failure rate 1.5%, 95%CI 0.4-3.7%) with an efficiency of 45% (95%CI 41-50%).

Conclusions: Combined with D-dimer testing, both "gestalt" and the Wells-rule are safe to rule-out PE in primary care. However, the Wells-rule is more efficient, and PE can be ruled-out in a larger proportion of suspected patients.

INTRODUCTION

Pulmonary embolism (PE) can be considered in patients with a wide variety of (pulmonary) symptoms, such as shortness of breath, coughing or pain on inspiration. Only in 10-30% of patients in whom PE is suspected, the diagnosis will be confirmed. (1) Thus, many diagnostic procedures are performed while PE is not present. To reduce the number of these redundant diagnostic procedures, guidelines recommend to start with identifying those patients at such low disease probability that referral or further diagnostics can safely be withheld. (2, 3) This risk stratification can either be based on an implicit physicians' estimate ("gestalt") or a formal diagnostic prediction model (like the Wells-rule or the (revised) Geneva score). (4-6) In patients identified as having low PE probability, a negative D-dimer test can safely rule-out PE subsequently. (7, 8)

Nowadays, formal prediction models are often regarded as a more accurate way to estimate disease probability compared to "gestalt". As it relies on pre-defined items, a prediction model is easy to use and results are independent of the level of experience. On the other hand, "gestalt" enables incorporation of individual characteristics, like the patient-specific context, that are not covered by prediction models. Many physicians regard that a standardized prediction model, sometimes even referred to as cookbook medicine, would not allow for such individual tailored diagnostics as much as "gestalt" does. (9)

The actual diagnostic performance of "gestalt" and prediction models in suspected PE has been compared in several studies in secondary care, yet with conflicting results due to substantial heterogeneity across studies. (10) For primary care however, evidence on the performance of "gestalt" in PE diagnosis is lacking altogether. The results from studies on "gestalt" performed in secondary care cannot directly be generalized to a primary care setting as family doctors (FDs) do not come across PE patients on a daily basis and thus inherently have less experience in recognizing PE cases as compared to hospital specialists. Moreover, hospital specialists often have access to some basic laboratory and imaging tests (e.g. blood gas analysis, chest x-ray, ECG, etc.) before a "gestalt" estimate is given. These tests are not readily available in primary care. Nevertheless, FDs do have much experience in distinguishing severe disease (like PE) from mild illnesses, often relying on the contextual knowledge of their patients, which is slowly constructed during the longstanding relationship that FDs have with most of their patients. (11)

Taking into account these possible merits and drawbacks of "gestalt" in primary care, as well as the mixed results regarding the comparison of "gestalt" versus a diagnostic decision rule from secondary care studies, the aim of this paper was to compare the diagnostic performance of "gestalt" and the Wells-rule to safely and efficiently rule out PE in a large primary care population with symptoms suggestive for PE.

METHODS

We used prospective data of the Dutch AMUSE-2 (Amsterdam, Maastricht, Utrecht Study on thrombo-Embolism) cohort. (7) This cohort was initially designed to evaluate the diagnostic performance of the Wells-rule combined with a qualitative D-dimer test in a Dutch primary care setting. A total of 662 consecutive adult patients, presenting at one of the 300 participating family doctors with symptoms raising suspicion of PE, were invited to participate. Sixty-four of these patients met one of the predefined exclusion criteria, leaving 598 patients for further evaluation. Further details of the cohort are described elsewhere. (7)

DIAGNOSTIC STRATEGIES

In all participants, FDs assessed relevant information on general health, specific cardiopulmonary or DVT signs and symptoms by systematically filling out a pre-specified case record form. FDs were asked to provide an estimated probability of PE being present, using a visual analogue scale with a range from 0% to 100% ("gestalt"). In addition, the seven items of the Wells-rule were scored. Finally, a qualitative point-of-care D-dimer test (Simplify D-dimer; Clearview, Inverness Medical, Bedford, UK) was performed in all patients.

OUTCOME ASSESSMENT

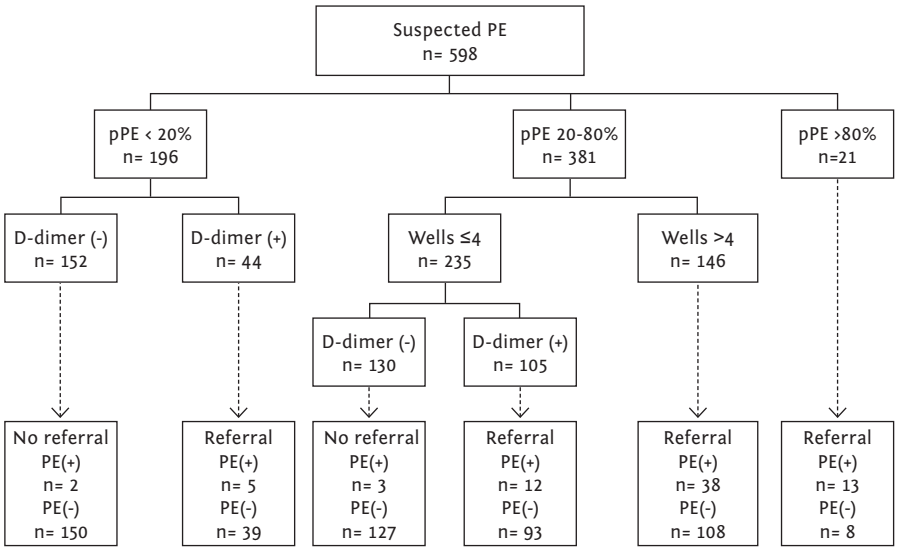
The study protocol required referral of all suspected PE patients to secondary care, and the standard diagnostic pathway according to current local hospital guidelines was followed. This usually comprised a combination of estimated probability and quantitative laboratory based D-dimer testing and if indicated a spiral computed tomography of the pulmonary arteries (CTPA). The primary outcome was presence or absence of venous thromboembolism, based on a combined reference standard of all diagnostic imaging tests performed in the hospital (spiral CT, ventilation-perfusion scanning, pulmonary angiography, leg ultrasonography and clinical probability assessment as performed in secondary care, with or without D-dimer testing) and including any occurrence of venous thromboembolic events during 3 months of follow-up in primary care. (7)

STATISTICAL ANALYSES

Some diagnostic variables had missing data, ranging from 0.5% for heart rate >100 beats/min, 2.7% for results of the D-dimer test and 16% for "gestalt". To minimize the effect of bias associated with selectively ignoring patients having missing data, we imputed all missing data using multiple imputation techniques prior to our current analyses. (12, 13)

The diagnostic performance of "gestalt" and the Wells-rule were compared using various methods. First, we quantified and compared the c-statistic (AUC; area under the curve) of the receiver operating characteristics (ROC) curve for "gestalt" and the Wells-rule, respectively. An AUC of 0.5 reflects no discriminative

FIGURE 1 FLOW CHART OF THE "STEPPED APPROACH" OF A FIRST PROBABILITY ASSESSMENT BASED ON "GESTALT" IN THE AMUSE-2 COHORT. FURTHER MANAGEMENT DEPENDS ON THIS ESTIMATED PROBABILITY OF PULMONARY EMBOLISM PRESENCE: REFERRAL IF PREDICTED RISK IS HIGH, NO REFERRAL IF PREDICTED RISK IS LOW



pPE= "gestalt", PE = pulmonary embolism.

ability, whereas an AUC of 1.0 indicates perfect discrimination.

Second, we stratified patients into groups of high or low predicted PE probability based on the "gestalt" or Wells-rule estimate, respectively, each combined with the point-of-care D-dimer test. For "gestalt", no common cutoff exists. We chose to apply a threshold for low probability of "gestalt" of ‹20% and a negative D-dimer test, in line with prior research in the field of venous thrombosis. (14, 15) In sensitivity analyses, we varied this threshold to 10% and 30%. The low-risk threshold used for the Wells-rule was a score ≤4 and a negative D-dimer test in accordance with previous publications. (1, 5) The efficiency (patients in the low PE probability category as proportion of total number of patients) and failure rate (proportion of patients with PE in this low probability category) of both strategies were calculated, next to sensitivity and specificity, and were compared.

Finally, we hypothesized that FDs using "gestalt" can correctly identify those patients at very low and very high estimated risk, but have difficulties who to refer in the "intermediate" estimated risk group. Therefore, we also assessed a combined approach, using "gestalt", the Wells-rule and D-dimer (see *Figure 1*). According to this "stepped approach", a D-dimer test is performed if the estimated "gestalt" probability is below 20%. Then, a negative D-dimer result safely rules out PE, whereas referral is indicated in case of a positive test result. In case of

TABLE 1 BASELINE CHARACTERISTICS OF THE AMUSE 2 COHORT, CONSISTING OF 598 PATIENTS SUSPECTED OF PULMONARY EMBOLISM IN PRIMARY CARE

	PE final diagnosis N=73	**Other diagnosis** N=525
Age (years), mean (SD)	53 (15)	47 (16)
Males, no. (%)	25 (34.2)	148 (28.2)
Clinical signs & symptoms of DVT, no. (%)	26 (35.6)	31 (5.9)
Alternative diagnosis less likely, no. (%)	61 (83.6)	272 (51.8)
Heart rate >100/ min, no. (%)	25 (34.2)	86 (16.4)
Immobilization, no. (%)	23 (31.5)	71 (13.5)
Previous PE or DVT, no. (%)	18 (24.7)	66 (12.6)
Hemoptysis, no. (%)	5 (6.8)	16 (3.0)
Malignancy, no. (%)	5 (6.8)	21 (4.0)
"gestalt" (estimated PE probability by FD), median % (IQR)	70 (40)	30 (32)
Wells score, median (IQR)	4.5 (4.0)	3.0 (3.5)
D-dimer Simplify positive, no. (%)	61 (83.6)	198 (37.7)

PE = pulmonary embolism; SD= standard deviation; no. = number; DVT = deep venous thrombosis; min= minute; FD= family doctor; IQR= inter quartile range

TABLE 2 CLINICAL DIAGNOSTIC PERFORMANCE MEASURES FOR THE WELLS-RULE ≤4 AND THE "GESTALT" ESTIMATE (LOW RISK CUTOFF <10%, <20% AND <30%) IN COMBINATION WITH D-DIMER TESTING, IN THE AMUSE-2 COHORT CONSISTING OF 598 PATIENTS SUSPECTED OF PULMONARY EMBOLISM IN PRIMARY CARE

Diagnostic parameter	**Wells ≤4 in combination with D-dimer**	**"Gestalt" <20% in combination with D-dimer**	**Sensitivity analysis 'gestalt' in combination with D-dimer**	
			"Gestalt" <10%	"Gestalt" <30%
Sensitivity	69/ 73 (95% (87-98%))	71/ 73 (97% (90-99%))	71/ 73 (97% (90-99%))	70/ 73 (96% (88-99%))
Specificity	268/ 525 (51% (47-55%))	150/ 525 (29% (25-33%))	42/ 525 (8% (6-11%))	216/525 (41% (37-45%))
Efficiency	272/ 598 (45% (42-50%))	152/ 598 (25% (22-29%))	44/ 598 (7% (6-10%))	219/ 598 (37% (33-41%))
Failure rate	4/ 272 (1.5% (0.6–3.7%))	2/ 152 (1.3% (0.4-4.7%))	2/ 44 (4.5% (1.3–15.1%))	3/ 219 (1.4% (0.5-3.9%))

95% confidence intervals are presented in between brackets.
Efficiency = proportion of low-risk patients in the total cohort (Wells ≤4, pPE <10%, 20% or 30%, and negative point-of-care D-dimer test) (i.e. patients not referred for objective testing).
Failure rate = the proportion of patients with pulmonary embolism in the low-risk group.

an estimated probability >80%, PE suspicion is so pronounced that the patient is referred immediately, without further procedures. In the remaining group with an estimated probability between 20 and 80%, the Wells-rule, followed by a D-dimer test in case of a score ≤4, is applied. We calculated the efficiency and failure rate for this "stepped approach" as well.

For all analyses, performed in IBM SPSS version 21, statistical significance was tested two-sided and defined as a p-value <0.05.

RESULTS

The cohort consisted of 598 patients with suspected PE, of whom 73 patients were diagnosed with venous thromboembolic disease: 72 pulmonary embolisms and 1 deep venous thrombosis events (prevalence 12%). Twenty-nine percent of all patients was male, and the mean age was 48 years. More baseline characteristics are presented in *Table 1*. The median "gestalt" estimate was 33% (interquartile range (IQR) 40%), with a total range from 0% up to 95%. Patients in whom the diagnosis PE was confirmed ultimately had a median estimated "gestalt" of 70% (IQR 40%), compared to 30% (IQR 32%) in those without PE (p=<0.001).

As can be seen in *Figure 2*, both "gestalt" and the Wells-rule had good overall discriminative ability to diagnose pulmonary embolism, with an area under the ROC curve of 0.77 (95% CI 0.70-0.83) and 0.80 (95% CI 0.75- 0.86) respectively.

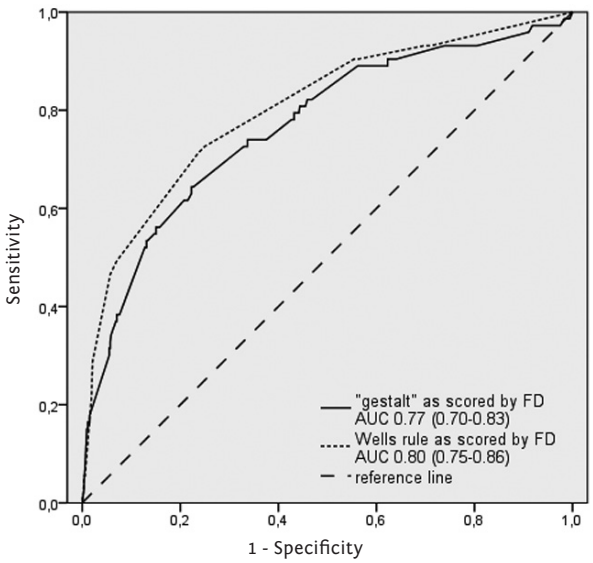
Using "gestalt" <20% plus a negative D-dimer test, 152 patients had low predicted PE probability (efficiency 25% (95% CI 22%-29%)) with a failure rate of 1.3% (95% CI 0.2%-4.7%) (see *Table 2*). A conservative threshold (<10%) in combination with a

TABLE 3 THE DIAGNOSTIC PERFORMANCE MEASURES FOR THE "STEPPED APPROACH", USING BOTH "GESTALT", D-DIMER AND THE WELLS-RULE, IN THE AMUSE-2 COHORT CONSISTING OF 598 PATIENTS SUSPECTED OF PULMONARY EMBOLISM IN PRIMARY CARE

	"Stepped approach"
Sensitivity	68/ 73 (93% (85-97%))
Specificity	277/ 525 (53% (48-57%))
Efficiency	282/ 598 (47% (43-51%))
Failure rate	5/ 282 (1.8% (0.7-4.1%))

Efficiency = proportion of low-risk patients in the total cohort (i.e. patients not referred for objective testing). 95% confidence intervals are presented in between brackets.
Failure rate = the proportion of patients with pulmonary embolism in the low-risk group.

FIGURE 2 RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE OF THE WELLS-RULE AND THE "GESTALT" ESTIMATE OF THE PROBABILITY OF HAVING PULMONARY EMBOLISM IN PRIMARY CARE, WITH THE CORRESPONDING AREAS UNDER THE CURVE (AUC) AND 95% CONFIDENCE INTERVALS IN THE AMUSE-2 COHORT CONSISTING OF 598 PATIENTS SUSPECTED OF PULMONARY EMBOLISM IN PRIMARY CARE



FD= family doctor

negative D-dimer was associated with very low efficiency (44/598 patients (7% (95% CI 5%-10%))) and 2 missed cases (failure rate 4.5% (95% CI 0.5%-15.5%)) (see *Table 2*).

Based on the Wells score of ≤4 plus a negative point-of-care D-dimer test, 272 patients were stratified into the low predicted PE probability group (efficiency 45% (95%CI 42-50%), of whom 4 patients were diagnosed with PE ultimately (failure rate 1.5% (95% CI 0.4%-3.7%)). The "stepped approach" combining "gestalt", the Wells-rule and D-dimer testing led to efficiency and failure rates comparable to the Wells-rule plus D-dimer without "gestalt" (see *Table 3* and *Figure 1*).

DISCUSSION

MAIN FINDINGS

In this study we compared the performance of "gestalt" and a formal diagnostic prediction model (i.e. the Wells-rule) to rule out PE in those suspected in primary care. Both diagnostic strategies had good overall discriminative ability with AUCs of 0.77 and 0.80. In combination with a qualitative D-dimer both "gestalt" and the

prediction model could safely rule-out PE. However, the number of patients that needs to be referred for objective testing was substantially lower when using the Wells-rule (efficiency 45% versus 25%), as well as for the "stepped approach".

COMPARISON WITH EXISTING LITERATURE

In 2011, Lucassen assessed the performance of "gestalt" and diagnostic prediction models for diagnosing pulmonary embolism in a meta-analysis. (10) There was substantial heterogeneity among all studies evaluating "gestalt". For instance, the thresholds for low probability used ranged from 10 and 40% for different studies. Furthermore, all studies were conducted in a secondary care setting. Our current findings however are in line with the meta-analysis' main conclusions: FDs do very well in safe exclusion of those patients at very low risk by combining "gestalt" with D-dimer testing, yet at the price of referring (much) more patients as compared to using a formal decision rule.

Barais et al. performed a qualitative study in a French primary care setting. (16) Using semi-structured interviews they aimed to unravel the process preceding a confirmed diagnosis of pulmonary embolism. For all interviewed FDs, the diagnostic process was mainly driven by intuitive factors, highlighting the importance of contextual knowledge and evidence in primary care. However, given the qualitative nature of this study, it does not give information if "gestalt" or a prediction model is more suitable to efficiently and safely identify low risk patients.

STRENGTHS & LIMITATIONS

To our knowledge, this is the first study on the diagnostic performance of both "gestalt" and a PE diagnostic prediction model in a primary care setting using a large cohort of patients suspected of PE.

However, some limitations need to be addressed for full appreciation of our findings. First, this is a post-hoc analysis of a prospective cohort study with the main aim to validate the Wells-rule in a primary care setting. As such, the estimated disease probability by "gestalt" and the Wells-rule were reported on the same case report form, introducing the chance of contamination of the estimates: "gestalt" might be influenced by the score of the Wells-rule. As a consequence, the results of "gestalt" and the Wells-rule are likely to be more alike, leading to dilution of the difference between the estimates. Nevertheless, distinct differences between the "gestalt" estimate and the Wells prediction were observed in this study, illustrated by the structurally overestimated "gestalt" estimate. As such, we hypothesize that the possible contamination only might have attenuated our current inferences and that the real life differences are expected to be larger in favor of the diagnostic prediction model.

Second, no consensus exists on the optimal threshold for low and high estimated PE probability. For this current analysis, we based the threshold on previously used thresholds in this field of research. (14) Furthermore, we evaluated the

performance using 3 different cutoff values to see the effect of the chosen cutoff on the diagnostic performance. Rather than absolute probabilities, this threshold should be seen as a reflection of the degree of uncertainty, often categorized as low, intermediate or high estimated probability.

Third, the Wells-rule is a structured diagnostic prediction model, yet comprises the subjective item "PE most likely diagnosis" with a relatively large contribution of 3 points into the score. (17) Therefore, it can be discussed that "gestalt" and the subjective item of the Wells-rule more or less "catch" the same information. The Wells-rule combines this information with other diagnostic variables, hence it may be expected to give better predictions when compared to analyzing only the "gestalt" information. However, the discriminative value of both strategies is comparable (AUC 0.77 (95% CI 0.70-0.83) vs 0.80 (95% CI 0.75-0.63)). Although this remains speculative, it seems that the objective items of the Wells-rule attenuates the (often overestimated) disease probability when using "gestalt" only. The integration of a subjective item with multiple (pre-selected) objective items may be pivotal in order to enhance its value for use in clinical practice. This is also exemplified from psychology research, which demonstrated that physicians incline towards overestimation of predicted risk, especially if a potentially severe diagnosis such as PE is considered. (11) In case of any doubt (that is: "intermediate risk"), physicians tend to be hesitant and thus – to be on the safe side – refer or initiate treatment relatively easily. This subsequently enhances over-diagnostics, over-treatment and increased health care use. (18, 19)

Finally, scoring of the "gestalt" estimate is likely to be influenced by levels of experience and individual style. Albeit speculative, experienced FDs may take prior experience into account and thus might feel more confident in assigning a low disease probability, though the opposite may also be true if this prior experience included a missed PE case. (16) Unfortunately, we do not have insight in the individual range of scoring probabilities and years of experience of all participating FDs. Over 300 FDs participated in this study, and most of these FDs included only a few patients each. We therefore expect random variation in scoring patterns between different FDs, minimizing its influence on the main inferences from our analyses. It would be interesting to study the impact of physicians' experience on assessing "gestalt" in primary care, yet the data of the current cohort unfortunately do not allow for such an evaluation.

CLINICAL IMPLICATIONS

Current guidelines recommend the use of a (structured) estimated disease probability to rule-out pulmonary embolism. Our current findings underline the use of a prediction model, but leave room for relying on "gestalt" if disease presence or absence is highly likely (or unlikely). As we expected, FDs seem very well capable to identify patients at both ends of the probability spectrum. However, for a large group of patients at intermediate risk, application of the Wells-rule and D-dimer testing will optimize the risk stratification better than using "gestalt" only.

CONCLUSIONS

Combined with D-dimer testing, both "gestalt" and the Wells-rule are safe to rule-out PE in this primary care cohort. However, as compared to intuitive diagnostic reasoning ("gestalt") the Wells-rule is more efficient, and PE can be ruled-out in a larger proportion of patients.

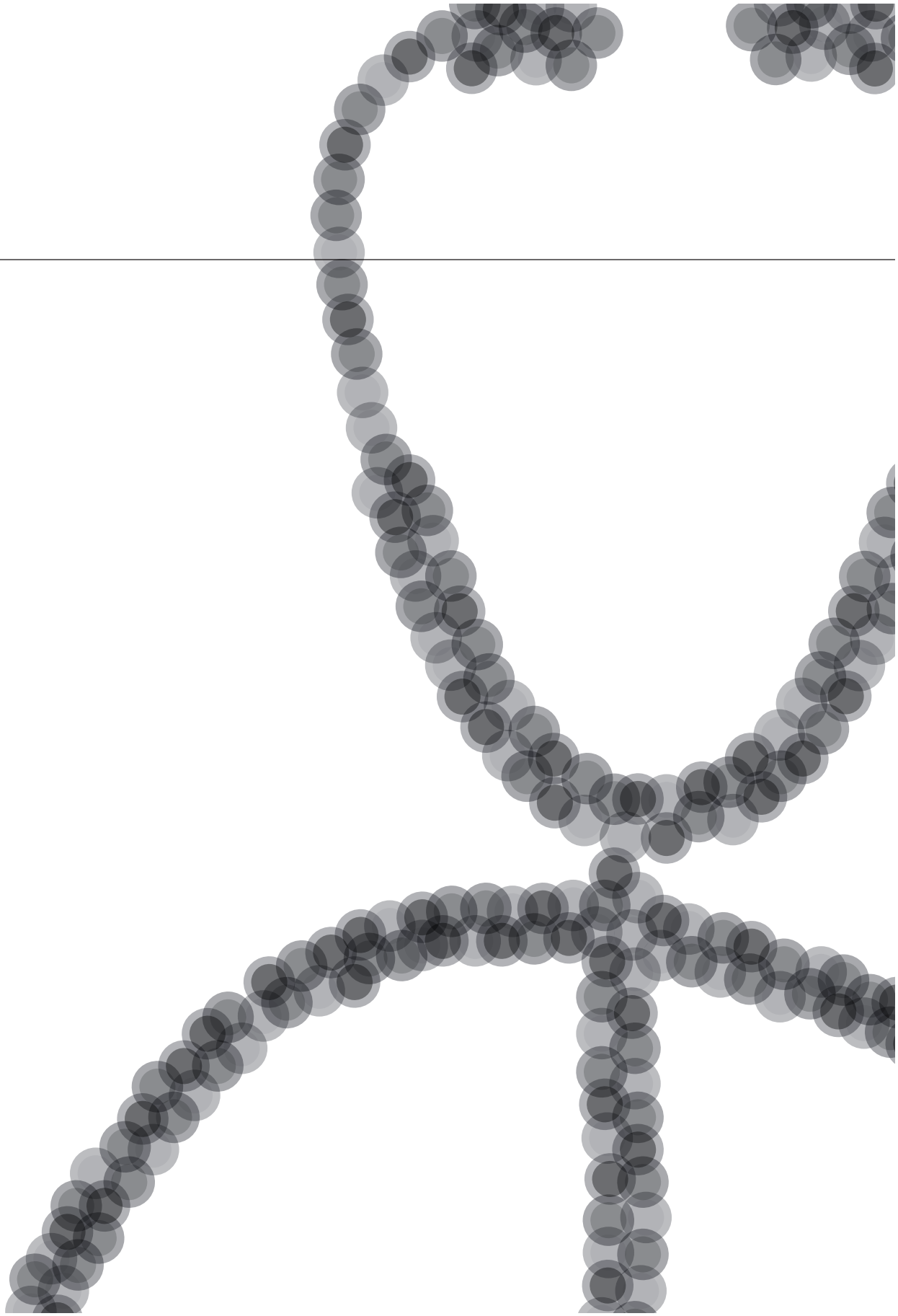
ACKNOWLEDGEMENTS

We thank AMUSE-2 project members Ruud Oudega, Hugo ten Cate and Martin Prins for their contribution to the design and initiation of the AMUSE 2 cohort.

REFERENCES

1. the PIOPED investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism: Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). JAMA. 1990;263(20):2753-2759.
2. Bates S, Jaeschke R, Stevens SM, et al. Diagnosis of DVT antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2):e351S-e418S.
3. Konstantinides SV, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Zompatori M. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. Eur Heart J. 2014;35(43):3033-3073.
4. Wells P, Andersen D, 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(3):416-420.
5. 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(3):165-171.
6. 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(1):92-97.
7. Geersing GJ, Erkens PM, Lucassen W, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
8. van Belle B, Buller H, Huisman M, Kaasjager K, Kamphuisen P. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. JAMA. 2006;295(2):172-179.
9. Cervellin G, Borghi L, Lippi G. Do clinicians decide relying primarily on Bayesians principles or on Gestalt perception? Some pearls and pitfalls of Gestalt perception in medicine. Intern Emerg Med. 2014;9(5):513-519.
10. Lucassen W, Geersing GJ, Erkens PM, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. Ann Intern Med. 2011;155(7):448-460.
11. Starfield BH. Contribution of primary care to health systems and health. Milbank Q. 2005;83(3):457-502.
12. Janssen KJ, Vergouwe Y, Donders AR, et al. Dealing with missing predictor values when applying clinical prediction models. ClinChem. 2009;55(5):994-1001.
13. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med. 1991;10(4):585-598.
14. Sanson BJ, Lijmer JG, Mac Gillavry MR, Turkstra F, Prins MH, Büller H. Comparison of a clinical probability estimate and two clinical models in patients with suspected pulmonary embolism. Thromb Haemost. 2000;83(2):199-203.
15. Geersing GJ, Janssen K, Oudega R, et al. Diagnostic classification in patients with suspected deep venous thrombosis: physicians' judgement or a decision rule? Br J Gen Pract. 2010;60(579):742-748.

16. Barais M, Morio N, Cuzon Breton A, et al. "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(5):e98112.
17. Klok FA, Karami Djurabi R, 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):1079-1080.
18. Harteis C, Billett S. Intuitive expertise: theories and empirical evidence. Educ Res Rev. 2013(9):145-157.
19. Grover SA, Lowensteyn I, Esrey KL, Steinert Y, Joseph L, Abrahamowicz M. Do doctors accurately assess coronary risk in their patients? Preliminary results of the coronary health assessment study. BMJ. 1995;310(6985):975-978.



CHAPTER 5

THE COST-EFFECTIVENESS OF POINT-OF-CARE D-DIMER TESTS COMPARED WITH A LABORATORY TEST TO RULE OUT DEEP VENOUS THROMBOSIS IN PRIMARY CARE

Janneke MT Hendriksen, Geert-Jan Geersing, Suzanne C van Voorthuizen, Ruud Oudega, Arina J ten Cate- Hoek, Manuela A Joore, Karel GM Moons, Hendrik Koffijberg

Expert Rev Mol Diagn. 2015;15(1):125–36.

ABSTRACT

Objective: Point-of-care (POC) D-dimer tests have been developed to exclude deep venous thrombosis quickly and on the spot, but are known to have lower sensitivity compared with laboratory-based tests. Their cost–effectiveness is still unknown.

Methods: We updated and extended a previously published Markov model to assess the cost–effectiveness of POC D-dimer tests ('Simplify', 'Cardiac', 'Triage' and 'Nycocard') compared with a laboratory-based latex assay to diagnose deep venous thrombosis in primary care.

Results: The 'Laboratory' strategy resulted in 6.986 quality-adjusted life years at the cost of €8354 per patient. All POC D-dimer tests resulted in health outcomes similar to the 'Laboratory' strategy. The 'Simplify' strategy maximized cost savings (–€155 [95% CI: –€246 to –€83]).

Conclusions: POC D-dimer tests yield similar health outcomes as laboratory-based testing procedures but can be performed more easily and at lower costs. Therefore, these tests are an alternative to laboratory-based testing and might be considered for exclusion of deep venous thrombosis in primary care.

INTRODUCTION

With an estimated annual incidence of 1 per 1,000 persons, deep venous thrombosis (DVT) is a common disease (1) and the third leading cause of vascular death after myocardial infarction and stroke, due to a potentially fatal pulmonary embolism. (2) Signs and symptoms of deep venous thrombosis are often non-specific and primary care physicians (PCPs) therefore have a low threshold of suspicion in patients with leg complaints. (3) As a consequence, the number of suspected DVT events is much higher than the number of confirmed cases: seven suspected to one confirmed event. (4) Hence, referral of all suspected patients to secondary care to undergo the reference standard – compression ultrasonography (CUS) – is undesirable as it unnecessarily increases patient burden and results in inefficient use of already limited healthcare resources.

The combination of a clinical decision rule (CDR) and D-dimer testing can help to mitigate this problem. DVT can safely be excluded in low-risk patients (i.e. those with a low estimated probability based on a CDR) with a negative D-dimer test. Such a rule-out strategy enables a safe exclusion rate of around 30% to 50% without referral to secondary care, as only high-risk patients are referred to the hospital for CUS. (5) In daily primary care practice, however, this still implies that low-risk patients are referred to a central lab facility to undergo D-dimer testing. This can lead to a delay in diagnosis and involves travel costs and a burden to patients. Notably in rural areas, but also in frail elderly, this is undesirable. Furthermore, laboratory based D-dimer tests take 40- 60 minutes to return test results, and often even longer before the result is returned to a primary care practice. All these issues can form a barrier for a successful implementation of a 'CDR and D-dimer' strategy in patients with suspected DVT, notably in primary care, where most suspected patients are actually managed.

Point-of-care, or "near patient", tests (POCTs) have been introduced as a potential solution for this problem. These POCTs can be performed during the patient consultation and yield results within 10-15 minutes. This ease and speed comes at a price: literally, as a single POCT is usually more costly than bulk analyses in a central lab facility, but also sensitivity of POCTs is slightly lower than sensitivity of laboratory tests. Potentially, this can lead to more 'missed' cases ('false negatives'). (6) Recent studies mainly focused on this sensitivity (or false negative rate) of POC versus laboratory-based D-dimer testing, as from a clinical point of view the number of potentially 'missed' cases is obviously of great importance. (7-17) Yet, with the continuously increasing strain on the healthcare budget, society at large also asks from physicians, laboratory staff, and policy makers to assess the full spectrum of all actual health outcomes, which also includes quality of life and costs of applied diagnostic strategies. (18, 19) Ideally, a randomized controlled diagnostic trial is needed to evaluate these health outcomes by comparing a POCT strategy with a laboratory-based strategy. Such a trial however is extremely costly as well as difficult to perform, and – not surprisingly – randomized controlled

diagnostic trials are therefore extremely scarce. An alternative approach – and aim of this study- comprises a formal model-based analysis comparing all costs and effects related to the use of different types of D-dimer (including POCT). To do so, we first performed a systematic literature review. In the review, we first wanted to identify all currently available D-dimer POCT (including a critical appraisal of identified studies). (20-22) Second, we aimed to assess the best estimate of diagnostic accuracy of these POCTs in the diagnosis of DVT in outpatients, to be used as input parameters for the model.

Next we assessed the cost-effectiveness of using either a POCT D-dimer or a laboratory-based assay as part of the primary care diagnostic work-up for deep venous thrombosis.

METHODS

DATA SOURCES AND SEARCHES

Identification of POC D-dimer tests. We identified D-dimer POCTs currently used for exclusion of DVT, based on a previously published diagnostic meta-analysis. (6) As new tests might have been introduced after publication of this meta-analysis, we repeated the systematic literature search using the PubMed database. We restricted our search to studies published in the English language. All identified papers were checked on potentially relevant other studies using their reference list. Finally, a clinical chemist, specialized in POCT, was consulted to assess if no relevant papers were missed. D-dimer POCTs differ in level of user-friendliness in a primary care setting. (23) For further data-extraction, we only considered tests that can easily be applied in primary care, that is a hand-held portable system or small analyzer, with ease of performance and appropriate analyzing speed.

Data extraction. We critically appraised all papers on D-dimer POCTs, identified by the updated literature search. From these papers, we extracted diagnostic accuracy measures (that is, sensitivity and specificity) and incorporated the best-available evidence in the diagnostic strategies (see below). The methods for the critical appraisal and data extraction are described in detail in the meta-analysis by Geersing et al. (6)

Defining diagnostic strategies. Subsequently, three types of diagnostic DVT work-up strategies were defined, based on application of the clinical decision rule or D-dimer test in primary or secondary care (see *Figure 1a*). Diagnostic accuracies of the different tests were adjusted based on their position in the diagnostic strategy (pathway) (see *Table 1a*). For the sake of clarity, we will refer to the primary care diagnostic strategy, including a Simplify POC test as "Simplify strategy", the primary care diagnostic strategy including the hospital-based laboratory testing as "laboratory strategy" and the strategy with referral to hospital for further

testing for all patients, regardless of the estimated risk, as "hospital strategy". In our analysis, the laboratory strategy was the reference strategy (referring low-risk patients only to a central lab facility for D-dimer testing). The incremental performance of the POC strategies was assessed relatively to this reference strategy to determine if POC tests could potentially replace laboratory testing, in terms of safety but also including all other health outcomes and costs (see below).

DECISION MODELING

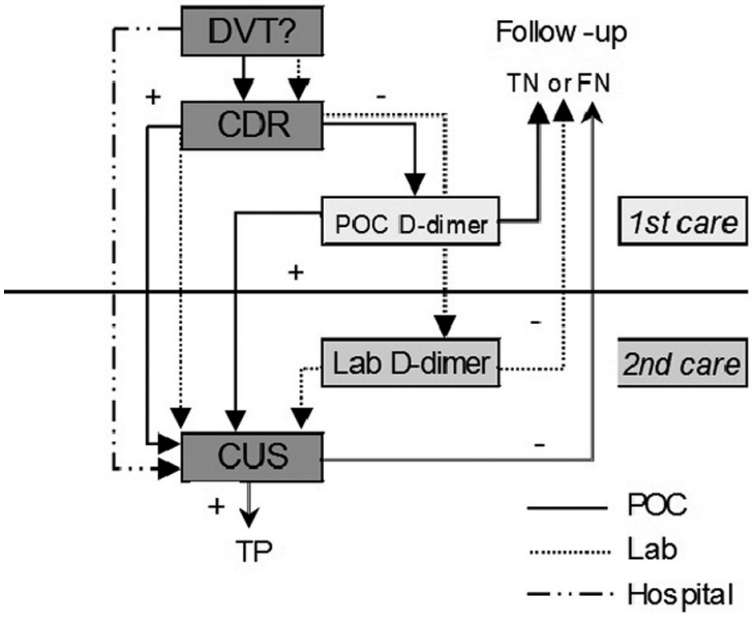
In case of scarce health care resources, choices have to be made on how to spend these resources on interventions and other diagnostic or therapeutic strategies in order to maximize health benefits for society. Comparing strategies then involves estimating their health benefits and costs. When evidence on health effects and costs is not collected directly, evidence synthesis and decision modelling can offer insight in both the health effects and costs of all possibly relevant alternative strategies, based on mathematical modelling. A commonly used method to gain insight in the health effects and costs of different diagnostic or therapeutic strategies is the construction of a Markov model. (24, 25)

MARKOV MODEL

As any mathematical model, a Markov model is a simplification of clinical reality. In a Markov model each patient has a health status that is tracked over time. This health status is reflected by a predefined number of (mutually exclusive) 'health states', and each patient is in one of these health states, at any point in time. Various events (e.g. occurrence of disease, treatment, complications from treatment, recovery etc.), can lead to transitions of patients between health states, with transitions taking place only in fixed time periods, so-called time cycles. Thus, in each cycle there is a certain probability of any patient of moving between health states. A key concept of the Markov model is that transition probabilities to new health states for each patient depend only on the current health state patients are in, that is, the model has no 'memory' of any previous health states the patient was in. In the model, each health state is associated with certain costs and certain quality of life (a utility value). Over a predefined number of cycles, all costs and utilities can be aggregated and total costs and effects can be estimated. (24, 25) In our applied model, different diagnostic strategies will then lead to different health benefits and costs. The structure of a Markov model is based on knowledge of current guidelines and diagnostic procedures, whereas the input parameters, such as transition probabilities, costs and utilities, are based on clinical studies, existing literature evidence, registries or expert opinion.

Quality-adjusted Life Years. To compare the overall health effects of different interventions on the quality and duration of life of individuals, even among various disease areas, the concept of Quality-adjusted Life Years (QALYs) has been developed. This concept combines both duration of life and quality of life, with 1

FIGURE 1A SCHEMATIC DIAGNOSTIC PROCESS OF SUSPECTED DEEP VENOUS THROMBOSIS



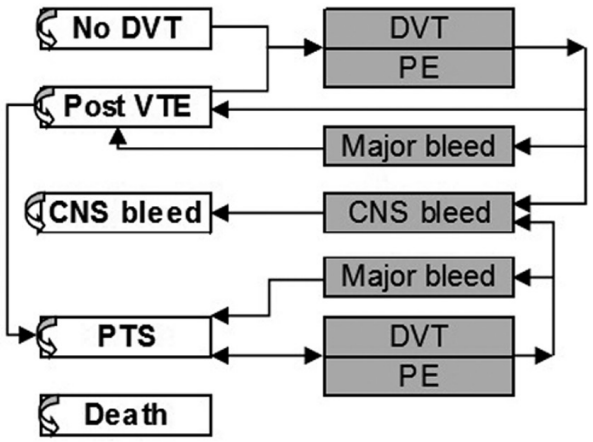
DVT? = suspected Deep Venous Thrombosis; CDR = clinical Decision Rule; CUS = compression Ultrasound; POC = point-of-care Strategy; Lab = laboratory Strategy; Hospital = hospital strategy; + = positive test result (either for CDR (high-risk), D-dimer testing or Compression Ultrasound (DVT present)); - = negative test result (either for CDR (low-risk), D-dimer testing or Compression Ultrasound (no DVT)); TP = true positives: CUS confirmed DVT patients, with high-risk CDR or low-risk CDR, but elevated D-dimer; TN = true negatives: No DVT present; FN = false negatives: DVT present but not detected by CUS in CDR high risk patients with, or in CDR low risk patients with positive D-dimer. Or DVT present in CDR low risk patients with negative D-dimer (no CUS performed).

Probability of referral for CUS: All high-risk patients based on CDR or low-risk CDR but elevated D-dimer.

QALY representing 1 full year of life lived in perfect health. In a Markov model, the number of QALYs for each intervention considered will depend on the total time (number of cycles) that patients spend in the health states that are part of the model, and on the utility assigned to these health states.

Costs. Costs are assigned to health states and events in the Markov model, and to the use of the considered interventions (diagnostic strategies). Different interventions may affect the time spend by patients in health states in different ways, for example by increasing the probability of correctly diagnosing disease, or reducing the risk of treatment complications. Consequently, the total costs of

FIGURE 1B MARKOV MODEL STRUCTURE



Health states are printed in white, events are colored grey. The final absorbing state is Death and can be entered from all health states and events.

VTE= venous thromboembolic event; DVT= deep venous thrombosis; PE= pulmonary embolism; CNS bleed= central nervous system bleeding; PTS= post-thrombotic syndrome.

an intervention or diagnostic strategy will vary, and differences in costs can be calculated between interventions or strategies.

Incremental Cost-effectiveness Ratio (ICER). By dividing the difference in costs of interventions by the difference in health benefits of interventions (QALYs) the Incremental Cost-Effectiveness Ratio (ICER) can be calculated. This ratio shows the balance of additional costs and health effects (positive or negative) for one intervention compared to another. When society is willing to set a threshold on the amount of money it is willing to pay for additional health gains, the ICER for interventions can be used to decide if new intervention are an acceptable use of scarce health care resources by investigating if its ICER falls below or above this threshold, called the ceiling ratio. In case of very small differences in effects between the different strategies, the estimates of ICERs can become unstable because effects of the different strategies are very similar. Then, cost-minimization can be applied to find the strategy with lowest cost.

TABLE 1A DIAGNOSTIC TEST AND STRATEGY PROPERTIES

Strategy properties**	Sensitivity of POCT ***	Sensitivity of strategy	p referral for CUS****	Costs (mean value (€))		Source
Point-of-care strategies				Variable$	269.46	
Triage	0.984 (0.947-0.998)	0.9228	0.6065	Fixed	82.27	(9, 19)
Cardiac	0.967 (0.920-0.993)	0.9134	0.4972	Fixed	67.19	(10-12, 19)
Nycocard	0.942 (0.882-0.981)	0.8989	0.6493	Fixed	53.54	(13, 14)
Simplify	0.915 (0.849-0.964)	0.8835	0.4989	Fixed	56.05	(4, 15-17, 19)
Laboratory strategy		0.9087	0.7533	Variable$	248.14	(7)
				Fixed	60.03	
Hospital strategy		0.9250	0.8417	Variable$	95.22	(7)
				Fixed	212.95	

POCT= point-of-care test; CUS= compression ultrasound.

** No standard error (SE) could be calculated directly for the strategies, as the mean relies on combined probabilities. Beta distributions were used for all separate parts of this combined data.

*** Sensitivity of 'best evidence' point-of-care test after correction for its use in patients with low-risk CDR.

**** Probability of referral for CUS: All high-risk patients based on CDR or low-risk CDR but elevated D-dimer.

$ Variable costs are dependent on referral for CUS. All costs in this table are rounded to cents.

5

MODEL DESCRIPTION

In order to assess cost-effectiveness of the POC strategies we extended and updated a previously constructed Markov model. (5) In short, this model simulates the course of events for a fictitious cohort of consecutive adult patients presenting at their PCP with suspected lower extremity DVT (at least swelling, redness or pain of the lower extremity) in primary care (see *Table 1b*). The characteristics of this cohort resemble the cohort of a previous prospective study (AmsterdamMaastrichtUtrecht Study on thromboEmbolism (AMUSE), 1002 consecutive adults with suspected lower extremity DVT, mean age 58 years, 37% males, 6% post-thrombotic syndrome and 17% recurrent thrombosis). (5) Diagnosis of DVT was confirmed in case of a non-compressible proximal vein on compression ultrasound according to the local hospital protocol, and rejected in case of 3-month uneventful follow-up.

The Markov model included five health states: *No DVT*, *Post VTE*, *Post-Thrombotic Syndrome (PTS)*, *Central Nervous System (CNS) bleed*, and *Death* (see *Figure 1b*). The health states No DVT and Post VTE account for the difference in risk of a first or recurrent VTE event. The health states PTS and CNS bleed, indicated by the white

TABLE 1B TRANSITION PROBABILITIES PER 6-MONTH CYCLE

Transition probabilities	Mean	SE	Distribution	Source
Prevalence of VTE and fatality				
VTE first cycle (= ratio event to complaint)	0.1357	0.0108	beta	(4)
VTE from No DVT after first cycle	0.0010	0.0001	beta	(1)
PE given VTE	0.2968	0.0017	normal*	(34)
Fatality of PE	0.4286	0.0700	beta	(35)
Recurrent VTE from Post VTE or PTS				
Till 6 months after VTE	0.0720	0.0064	beta	(36)
From 6 months till 1st year after VTE	0.0409	0.0049	beta	(36)
From 1st till 10rd year after VTE	gradual decline			(36)
PTS from Post VTE				
Till 1st year after VTE	0.0906	0.0152	dirichlet	(37)
From 1st till 2nd year after VTE	0.0338	0.0096	dirichlet	(37)
From 2nd till 10th year after VTE	gradual decline			(37)
Bleeding during VKA treatment				
Non-fatal, non-CNS major bleeding	0.0304	0.0011	dirichlet	(38)
Fatal major bleeding	0.0110	0.0007	dirichlet	(38)
Non-fatal CNS bleeding	0.0030	0.0004	dirichlet	(38)

VTE= venous thromboembolism; DVT= deep venous thrombosis; PE= pulmonary embolism; PTS= post-thrombotic syndrome; CNS= central nervous system; SE = standard error.

* A normal distribution instead of a beta distribution was used in MS Excel due to the high number of patients.

boxes, represent the persistent, life-long impact of a nonfatal intracranial bleeding or the post-thrombotic syndrome on quality-of-life and costs The events DVT, PE (Pulmonary Embolism), major bleed and CNS bleed, represented by gray boxes, are transient states in which individual can only stay for exactly one time cycle. The time cycle length (the size of all time periods in the model) was set to 6 months. The absorbing state Death can be reached from any state but arrows were omitted for visual clarity.

We chose a 10-year time horizon for the analysis, to assess the long-term health effects and costs of the different strategies, such as a recurrent event or PTS. The intended duration of anticoagulant treatment after a thrombotic event depends on the indication and ranges from 3 months to lifelong. Here, we assumed a fixed treatment period of 6 months regardless the type of event (DVT or PE). This resembled the AMUSE cohort best, in which the majority of confirmed DVT cases presented with a first episode of DVT and received 6 months of treatment. Only the events PE or DVT could lead to the bleeding states (a complication of anticoagulant treatment). The event and health state CNS bleed were assumed to

have similar utility. PTS predominantly develops within the first years after a VTE event. (26, 27) Therefore, we assumed that no new cases of PTS would arise beyond 5 years. As treatment for both PE and DVT is similar and will generally not be altered in case a PE develops after initiation for DVT treatment, we assumed that no PE would occur once the diagnosis DVT had been established by ultrasound. Treatment of severe hemodynamically instable PE with thrombolytic therapy and chronic thromboembolic pulmonary hypertension (CTEPH), a rare complication of PE, were not modeled because these complications are very rare and did not occur in the AMUSE cohort. We did account partially however for this potential limitation, by increasing the case-fatality rate of PE (to as high as 60%) in a sensitivity analyses. Transition probabilities were based on the AMUSE study and literature (See *Table 1b*). Risks for PTS and VTE recurrence were time dependent as these risks decrease over time. Background mortality was based on age- and sex-dependent 2010 mortality rates from Statistics Netherlands. (28)

Age-dependent norm utilities (i.e. quality of life scores) from the general population were used for *No DVT* and *Post VTE*. (29) Utilities for the health state *PTS* and the event *DVT* were derived from AMUSE EQ-5D assessments (5), whereas utilities for the remaining health states and events were obtained from the literature. (5, 26)

All costs, including those used in the former AMUSE analysis, were recalculated to 2010 costs using price index values from Statistics Netherlands. (31) In a sensitivity analysis, the base-case analysis was repeated using 2013 costs. Point-of-care D-dimer costs were retrieved from the manufacturer or distributor. The costs of one test include the purchasing costs, the analyzer and a periodical control test. An estimated 50 tests are performed per year per group practice consisting of 4 PCP's, based on incidence rates of (suspected) DVT. (1) Remaining costs were based on reference standard costs from the Dutch Costing Manual (32), on data derived from the Dutch Drug Information System of the Health Care Insurance Board, and expert opinion. Details on cost estimates are given in *Table 2 a-d*. Future costs and health effects were discounted with a yearly discount rate of 4% and 1.5% respectively, according to Dutch guidelines. (34)

BASE CASE ANALYSIS

The base case analysis compares the costs and effects of the presented diagnostic strategies. We then calculated incremental cost-effectiveness ratio's (ICERs) by dividing the incremental costs by the incremental Quality-adjusted Life Years (QALYs) for the POCTs and hospital based strategies compared to the laboratory reference strategy. Strategies were rated by the health benefits they provided and the strategy providing most health benefits at acceptable costs was identified. A probabilistic sensitivity analysis (PSA) was performed using Monte Carlo simulation with 5000 samples to account for uncertainty in all parameter estimates. Results were presented in an incremental cost-effectiveness plane to visualize uncertainty in cost-effectiveness estimates. (24, 35)

TABLE 2 DETAILS OF COST CALCULATIONS (EUROS, 2010 PRICE LEVELS)

TABLE 2 A COSTS OF DIAGNOSTIC STRATEGIES

	Costs per unit (€)	Resource use	Source
Point-of-care strategy			
At PCP (fixed)			
PCP consultation	28.36		(34)
# PCP consultations		1.00	Assumption
D-dimer test:			
Triage	39.73		Company market price
Cardiac	24.66		Company market price
Nycocard	11.01		Company market price
Simplify	13.52		Company market price
Performing D-dimer test	14.18		(34)
In hospital (variable)			
Blood draw	12.78		(32)
Lab procedure	2.13		(32)
# Lab procedures		4.00	Expert opinion
ER visit	152.92		(34)
Ultrasound	56.81		(32)
# Ultrasounds		1.67	(33)
Laboratory strategy			
At PCP and laboratory facility (fixed)			
PCP consultation	28.36		(34)
# PCP consultations		1.00	Assumption
Blood draw	12.78		(32)
D-dimer test	10.36		(32)
# Lab procedures		4.00	Expert opinion
Lab procedure	2.13		(32)
In hospital (variable)			
ER visit	152.92		(34)
Ultrasound	56.91		(32)
# Ultrasounds		1.67	(43)
Hospital strategy			
At PCP and in hospital (fixed)			
PCP consultation	28.36		(34)
# PCP consultations		1.00	Assumption
Blood draw	12.78		(32)
D-dimer test	10.36		(32)
Lab procedure	2.13		(32)
ER visit	152.92		(34)
In hospital (variable)			
Ultrasound	56.91		(32)
# Ultrasounds		1.67	(43)

PCP= primary care physician; ER= emergency room.

TABLE 2 B TRAVEL COSTS (TWO- WAY TRIP)

	Unit costs (2010 (€))	Resource use	Source
Travel to PCP			
Proportion to PCP by car		50%*	(4)
Km to PCP	1.1		(34)
Costs per km per car	0.20		(34)
Car parking	3.04		(34)
Total costs	3.27		
Travel to Hospital			
Proportion to hospital by car		100%	(4)
Km to PCP	7.0		(34)
Costs per km per car	0.20		(34)
Car parking	3.04		(34)
Total costs	8.91		

*For the proportion of patients that went to the PCP by car a beta distribution was used with a standard error of 0.02.

PCP= primary care physician; Km= kilometer

TABLE 2 C COSTS ASSOCIATED WITH HEALTH STATES

	Costs per unit (€)	Lower – upper values	Distribution	Source
Health states				
No DVT	0.00	-	Fixed	Assumption
Post VTE	0.00	-	Fixed	Assumption
PTS	3,564.17	153.71 - 11,614.09	Beta PERT*	(4, 44)
CNS bleed	31,195.69	-	Fixed	(45)
Transition				
PTS	3,696.46	300.63 – 11,713.00	Beta PERT*	(4, 44)

* The beta PERT (Program/ Project Evaluation and Review Technique) distribution is a version of the beta distribution which can be used when a minimum and maximum are available.

DVT= deep venous thrombosis; VTE= venous thromboembolism; PTS= post-thrombotic syndrome; CNS= central nervous system

TABLE 2 D COSTS ASSOCIATED WITH EVENTS

	Costs per unit (€)	Resource use	SE	Distribution	Source
Event DVT					
PCP consultation	28.36			fixed	(34)
# of PCP consultations		1.83	0.30	gamma	(4)
Home care compression therapy	526.61			fixed	(4, 34)
LMWH 7 days	82.97			fixed	(34)
Coumarins 6 months	117.12			fixed	(34)
specialist visit	64.81			fixed	(34)
# control visits specialist		2.79	0.84	gamma	(4)
INR control visit	9.75			fixed	Thrombosis Service
# INR control visits		16.38	1.28	gamma	(4)
compression stockings	68.53			fixed	Health Insurance company
Hospital day	582.31			fixed	(34)
# hospital days		0.63	0.11	gamma	(4)
Total costs	**1,555.56**				
Event PE					
PCP consultation	28.36			fixed	(34)
# of PCP consultations		2.42	1.07	gamma	(4)
ER visit	152.92			fixed	(34)
CT thorax	144.19			fixed	(34)
ECG	27.68			fixed	(34)
Blood draw	12.78			fixed	(32)
Lab procedures	2.13			fixed	(32)
# lab procedures		4.00		fixed	Expert opinion
Hospital day	582.31			fixed	(34)
# hospital days		7.00		fixed	Expert opinion
LMWH 7 days	82.97			fixed	(34)
Coumarins 6 months	117.12			fixed	(34)
# control visits specialist		2.79	0.84	gamma	(4)
# INR control visits		16.38	1.28	gamma	(4)
Total costs	**5,041.83**				
Event Major bleed	4,629.65	1,856.47- 12,641.05		Beta PERT*	Expert opinion**
Event CNS bleed	13,404.13			Fixed	(45)

DVT= deep venous thrombosis; PCP= primary care physician; LMWH= Low Molecular Weight Heparin; INR= international normalized ratio; PE= pulmonary embolism; ER= emergency room; CT= computed tomography; ECG= electrocardiography; CNS= central nervous system; SE= standard error.

* The beta PERT (Program/ Project Evaluation and Review Technique) distribution is a version of the beta distribution which can be used when a minimum and maximum are available. ** An extensive expert opinion evaluation of 4 experts (2 hematologists, a vascular internist and a gastroenterologist) was performed previously to assess the costs associated with a major (gastrointestinal) bleed.

SCENARIO ANALYSES

To assess the robustness of our results we repeated our analysis in a number of scenarios based on alternative sets of input values for the model. First, pooled diagnostic accuracy measures of all D-dimer tests instead of one single measure derived from literature were used. Second, we repeated our analysis based on an ELFA instead of latex assay as laboratory test. Sensitivity of ELFA is marginally higher than the sensitivity of latex assays (7, 8), both techniques are used as conventional D-dimers test with choice predominantly based on laboratory preference. A point-of-care analyzer is an expensive investment but may be worthwhile if several primary care physicians share these costs. In a third scenario we therefore lowered the number of assays performed per general practice per year to account for acquisition costs. Furthermore, we assessed a fourth scenario mimicking a rural rather than an urbanized setting by changing the mean travel distance to the hospital from 7 to 100 kilometers. Fifth, we lowered the case fatality rate of pulmonary embolism from 43% to 16% in a fifth scenario to assess its impact on the cost-effectiveness of the different tests in relating to missing a DVT. (36, 37) Then, we increased the case-fatality rate to 60% to account for even more complications due to either missing a DVT (such as CTEPH) or PE treatment (e.g. thrombolysis). Finally, the base-case analysis was repeated using 2013 costs.

RESULTS

SYSTEMATIC REVIEW OF APPLIED POINT OF CARE D-DIMER TESTS

The systematic search in PubMed yielded 1,676 papers. After screening of title and abstract we identified 62 POCT papers evaluating 71 POCTs. Of these papers, several Nycocard papers were excluded from further analyses due to a recent change of manufacturer and test characteristics. Furthermore, we did not take into account the SimpliRED D-dimer test, as it has been largely replaced by the Simplify test, and two other devices because of the relative large size of analyzer. Our final selection consisted of a total 13 POCT-analysis in 11 papers describing the diagnostic accuracy of four point-of-care devices that met our user-friendliness criteria: three quantitative (Triage (Biosite) (N=2) (9, 23), Cardiac (Roche) (N=4) (10-12, 23), Nycocard (Axis-Shield) (N=2) (13, 14) and 1 qualitative test (Clearview Simplify (Inverness Medical Innovations) (N=5) (4, 15-17, 23). The flowchart is provided in Appendix 1 and details of the four POCTs are provided in Appendix 2 and 3.

As can be seen in Appendix 2, most of the studies were conducted in secondary care, with subsequent higher DVT-prevalence that might influence the diagnostic accuracy measures. Furthermore, there was heterogeneity in terms of use of a CDR or not, the reference standard and cut-off value used. In total, we included data of 678 patients tested with Triage, 1,302 patients with Cardiac, 510 patients (Nycocard) and 2,334 patients tested with Simplify. Respectively 82/678 (12.1%), 361/1,302 (27.7%), 129/510 (25.3%) and 378/2,334 (16.2%) of these patients were diagnosed with venous

thrombosis. Since the heterogeneity between studies, we used the best-available evidence (that is: the study per test that best resembled our population) in the base-case analysis and all evidence (that is: pooled diagnostic accuracy estimates) in the sensitivity analysis. Low-risk population corrected single test- and complete diagnostic strategy sensitivity were respectively 0.984 and 0.923 (Triage), 0.967 and 0.913 (Cardiac), 0.942 and 0.899 (Nycocard) and 0.915 and 0.884 (Simplify). The complete diagnostic strategy sensitivity of referral to secondary care yielded a sensitivity of 0.909 for the Laboratory strategy, and 0.925 for the Hospital strategy.

BASE CASE ANALYSIS

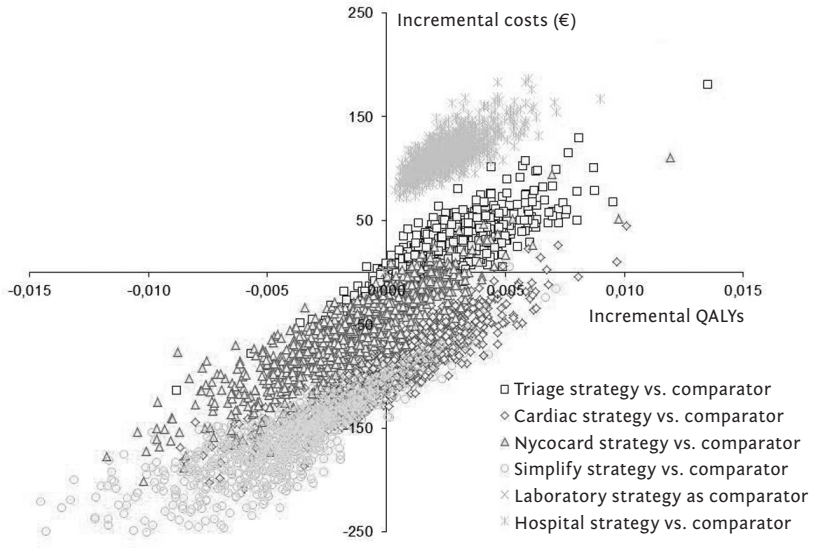
The laboratory strategy was the reference strategy and resulted in 6.986 QALYs at the cost of €8,354 per patient over the 10-year horizon. All other strategies resulted in comparable health outcomes although there were very small differences (see *Table 3*). The Nycocard and Simplify strategies resulted in marginally less QALYs, -0.0015 (95%CI -0.0074 to 0.0033) and -0.0040 (95%CI -0.0113 to 0.0015) respectively. The Cardiac, Triage and Hospital strategies resulted in fractional more QALYs, +0.0007 (95%CI -0.0041 to 0.0049), +0.0022 (95%CI -0.0014 to 0.0062) and +0.0025 (95%CI 0.0008 to 0.0054) respectively (see *Figure 2*). Because of the marginal differences in health outcomes and corresponding wide confidence intervals we chose to focus on cost minimization instead of cost-effectiveness and ICERs. The Simplify strategy was the least expensive strategy (- €155 (95% CI €-246 to €-83) per patient compared to the laboratory strategy, and for the hospital strategy (i.e. referring all patients directly for D-dimer testing to the hospital, regardless of an initial risk assessment) associated costs were highest (+€114 (95% CI €85 to €155)). In 96.3% of all simulations, the Simplify strategy appeared to be the least expensive diagnostic strategy. The Cardiac and Nycocard strategies were least expensive in

TABLE 3 COMPARISON OF COSTS AND EFFECTS OF DIFFERENT STRATEGIES WITH THE REFERENCE STRATEGY (LABORATORY)

	Costs	QALYs
Simplify	€-155.37 (95% CI €-245.83 to €-82.58)	-0.0040 (95% CI -0.0113 to 0.0015)
Nycocard	€-56.43 (95% CI €-136.32 to €10.20)	-0.0015 (95% CI -0.0074 to 0.0033)
Cardiac	€-83.20 (95% CI €-146.08 to €-21.66)	0.0007 (95% CI -0.0041 to 0.0049)
Triage	€16.87 (95% CI €-34.57 to €-75.50)	0.0022 (95% CI -0.0014 to 0.0062)
Laboratory	Reference	
Hospital	€113.59 (95% CI €85.10 to €155.35)	0.0025 (95% CI 0.0008 to 0.0054)

95%CI= 95% confidence interval; QALYs= quality-adjusted life years

FIGURE 2 COST-EFFECTIVENESS PLANE



QALY= quality-adjusted life years

2.3% and 1.4% of the simulations respectively, and all other strategies never were least expensive.

5

SCENARIO ANALYSES

In all of our predefined scenarios the analysis resulted in comparable outcomes, with only marginal differences in health outcomes between the different strategies. A lower case fatality rate for pulmonary embolisms (from 43% to 16%) did not lead to a significant shift in health outcome estimates or costs: In 96.5% of all simulations the Simplify strategy was the least expensive strategy, the Cardiac in 2.6% and the Nycocard in only 0.9%. A higher case fatality rate (60% instead of 43%) lowered the percentage of simulations in which the Simplify was the least expensive strategy to only 95.4%. Using costs recalculated to 2013 again did not change the order of these strategies as well.

DISCUSSION

We assessed the long-term cost-effectiveness of four different point-of-care D-dimer tests compared with a laboratory based D-dimer strategy in the diagnostic work-up of patients with suspected DVT in primary care. Compared to the laboratory strategy, the hospital-based strategy (that is: referral of all patients to secondary

care to perform D-dimer test, not only those with low risk) offers no substantial extra health benefits for the extra costs incurred. If the main interest is a reduction of costs, the Simplify strategy may be preferred, yet at the expense of marginal health loss that might come down to a DVT diagnosis missed. The robustness of these results were confirmed in several scenario analyses.

Our findings are in concurrence with results from the Markov model used in the first AMUSE study, demonstrating that a diagnostic strategy with a clinical decision rule and a qualitative POC D-dimer test can reduce the number of referrals for objective testing in secondary care by as much as 50% at the expense of missing only 1.4% of DVT cases (no fatal events). (4) Similar as to our findings, the limited amount of health loss associated with these missed non-fatal VTE cases was counterbalanced by substantial cost-savings and the primary care strategy therefore was considered cost-effective. In the current study, we extended this Markov model, updating all model parameters on POCT and DVT treatment with the results of an updated systematic literature review. Furthermore, we modeled the use of all different available POC and laboratory based D-dimer tests as part of a primary or hospital care diagnostic strategy.

Strengths of our study include the complete review of the point-of-care D-dimer tests available, and the ability to use data of a primary care management study on the accuracy of a POC test, performed by over 300 primary care physicians across the Netherlands. (4) Furthermore, we were able to incorporate additional time required in the PCPs office to examine both the quantitative and qualitative assays (up to 10 minutes). In addition, we looked into potential safety concerns related to POC testing by varying the case fatality rate of pulmonary embolisms in patients with missed DVT in a scenario analysis. The actual case fatality rate of untreated PE is unclear, thus we chose a rather high estimate of 43% for our base-case analysis to prevent too optimistic estimates, and even a 60% estimate in a scenario analysis. This way, we also tried to account for a scarce CTEPH event or treatment complications of severe PE as our data lacked details of these events. Even with these rather high estimates, (less sensitive) POCTs still yielded health outcomes comparable to the more sensitive laboratory based tests and are thus worth considering from a health economic point of view.

Our study also had certain limitations. First, the model and input parameters are based on the current Dutch health-care system and its underlying cost structure. As such, this may lessen the generalizability of our findings to other countries. However, with the use of real primary care data we believe that results are particularly transferable to health care settings with a strong primary care background, as is the case in large parts of Europe, Oceania and parts of Northern-America.

Second, we did not distinguish between quantitative and qualitative D-dimer

tests other than with respect to diagnostic accuracy and costs. In practice, differences in user-friendliness of POCTs may influence their potential for use in daily practice. Although qualitative tests may be preferred in daily practice given that 'only' a finger prick is needed for these tests (instead of a venipuncture that is needed for the quantitative tests), previous studies have demonstrated that inter-observer variability of interpreting qualitative D-dimer tests (Simplify) can result in an unacceptable decrease of test sensitivity (more 'false-negatives'), especially when performed by inexperienced users. (23) Yet, it is currently not possible to objectively include such (dis)advantages of the type of POC D-dimer test in the analysis. We would like to stress here however that implementation of POC D-dimer testing should always include a strong involvement of clinical chemists and local laboratories, in order to maintain quality assurance as well as training for use in daily clinical care.

Third, only fractional differences were found on health outcomes between POCTs and laboratory based D-dimer tests: for example, a difference of 0.0022 QALYs means living 0.0022*365 = 0.8 day = 19 hours, extra in perfect health. Although we acknowledge that the differences reflect the test characteristics and should not easily be bypassed, we believed that the differences found were too small to proceed actually perform a cost-effectiveness analysis. Instead, we focused on cost-minimization, thus providing good insight in the costs associated with different diagnostic strategies given that health effects are comparable. The small differences in QALYs between the strategies can be explained by the fact that the differences in sensitivity and specificity between all D-dimer testing methods are relatively small. Furthermore, the D-dimer test is part of diagnostic pathway in which other factors (e.g. accuracy of CDR) are of influence as well.

Finally, from a clinical standpoint the costs and burden of a false positive test result are obviously far less important than those of a false negative test result, that is, missing a potentially fatal venous thrombosis event. This would suggest that, at least for clinicians, optimal sensitivity and safety would be all that matters. However, money that is spent on one test, or one patient cannot be spent on other tests or patient with other diseases. Given the limited health care budget and increasing health care insurance premiums society therefore increasingly demands to also incorporate costs in medical-decision making. In our analysis, the marginal health loss induced by the use of some POC tests (caused by a minimal increase of false negatives) is largely outweighed by the reduced number of referrals, and thus fewer costs.

CONCLUSIONS

Although POC D-dimer tests are known to be slightly less sensitive than laboratory-based D-dimer tests they are able to limit the additional burden and costs associated with referral and additional testing (e.g. by CUS). Balancing these

aspects in our economic evaluation we found that the use of quantitative- and even qualitative- D-dimer POCTs as part of the primary care diagnostic work-up strategy for suspected deep venous thrombosis is likely to yield similar health outcomes as referral for a laboratory based D-dimer test in secondary care, but more rapidly and at lower costs.

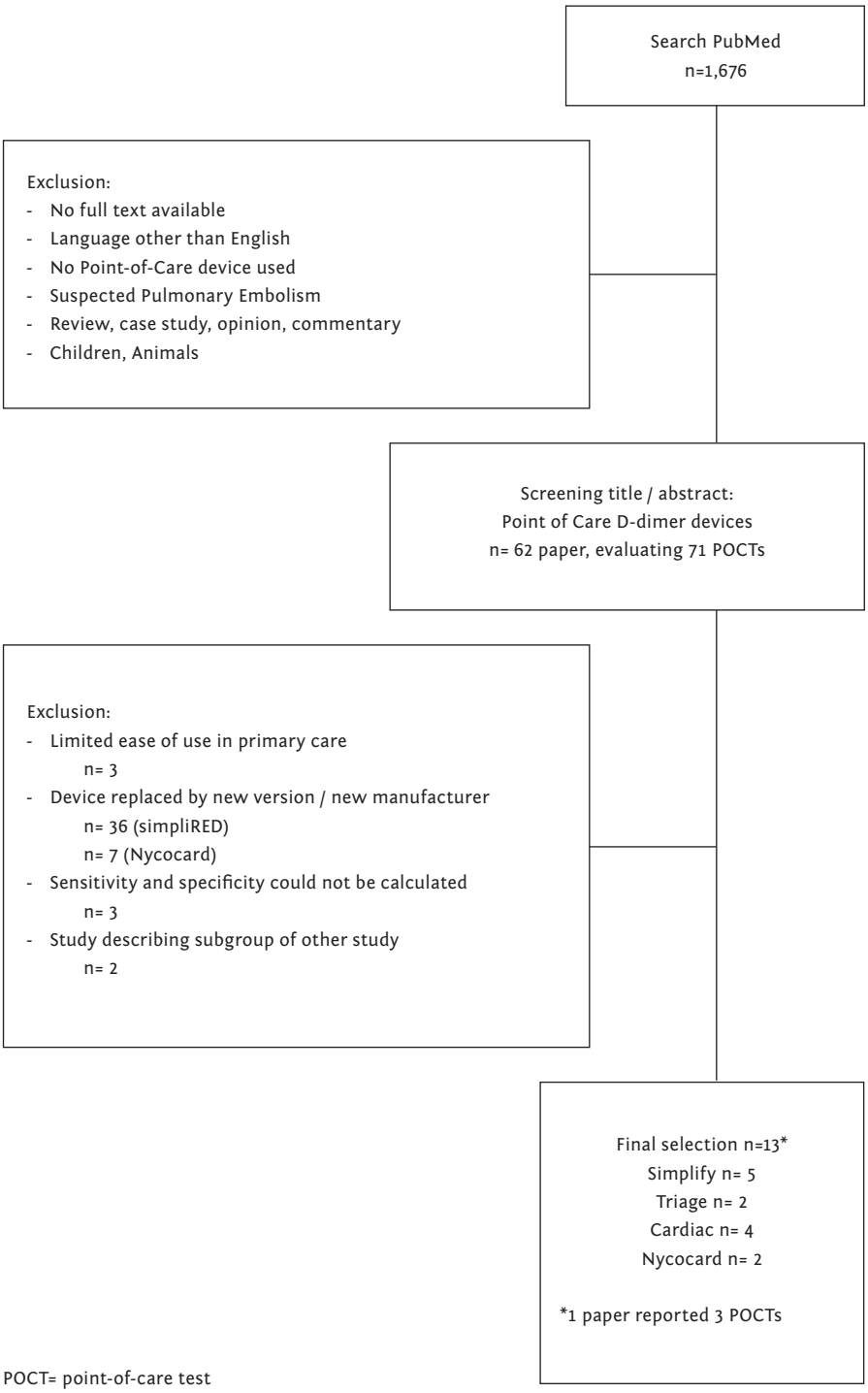
REFERENCES

1. White RH. The epidemiology of venous thromboembolism. Circulation. 2003;107(90231):I-4-8.
2. Goldhaber SZ. Preventing pulmonary embolism and deep vein thrombosis: a 'call to action' for vascular medicine specialists. J Thromb Haemost. 2007;5(8):1607-9.
3. Goodacre S, Sutton AJ, Sampson FC. Meta-analysis: the value of clinical assessment in the diagnosis of deep venous thrombosis. Ann Intern Med. 2005;143(2):129-39.
4. Buller HR, ten Cate-Hoek AJ, Hoes AW, Joore MA, Moons KGM, Oudega R, et al. Safely ruling out deep venous thrombosis in primary care. Ann Intern Med. 2009;150(4):229-35.
5. Ten Cate-Hoek AJ, Toll DB, Buller HR, Hoes AW, Moons KG, Oudega R, et al. Cost-effectiveness of ruling out deep venous thrombosis in primary care versus care as usual. J Thromb Haemost. 2009;7(12):2042-9.
6. Geersing GJ, Janssen KJM, Oudega R, Bax L, Hoes AW, Reitsma JB, et al. Excluding venous thromboembolism using point of care D-dimer tests in outpatients: a diagnostic meta-analysis. BMJ. 2009;339(b2880).
7. Di Nisio M, Squizzato A, Rutjes AWS, BÜLler HR, Zwinderman AH, Bossuyt PMM. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. J Thromb Haemost. 2007;5(2):296.
8. Stein PD, Hull RD, Patel KC, Olson RE, Ghali WA, Brant R, et al. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. Ann Intern Med. 2004;140(8):589-602.
9. Baker PM, Howgate SJ, Atherton J, Keeling DM. Comparison of a point of care device against current laboratory methodology using citrated and EDTA samples for the determination of D-dimers in the exclusion of proximal deep vein thrombosis. Int J Lab Hematol. 2010;32(5):477.
10. Legnani C, Fariselli S, Cini M, Oca G, Abate C, Palareti G. A new rapid bedside assay for quantitative testing of D-dimer (Cardiac D-dimer) in the diagnostic work-up for deep vein thrombosis. Thromb Res. 2003;111(3):149-53.
11. Bucek RA, Quehenberger P, Feliks I, Handler S, Reiter M, Minar E. Results of a new rapid D-dimer assay (Cardiac D-dimer) in the diagnosis of deep vein thrombosis. Thromb Res. 2001;103(1):17.
12. Dempfle CE, Korte W, Schwab M, Zerback R, Huisman MV. Sensitivity and specificity of a quantitative point of care D-dimer assay using heparinized whole blood, in patients with clinically suspected deep vein thrombosis. Thromb Haemost. 2006;96(1):79-83.
13. Gardiner C, Pennaneac'h C, Walford C, Machin SJ, Mackie IJ. An evaluation of rapid D-dimer assays for the exclusion of deep vein thrombosis. Br J Haematol. 2005;128(6):842-8.
14. Larsen TB, Stoffersen E, Christensen CS, Laursen B. Validity of D-dimer tests in the diagnosis of deep vein thrombosis: a prospective comparative study of three quantitative assays. J Intern Med. 2002;252(1):36-40.
15. Cini M, Legnani C, Cavallaroni K, Bettini F, Palareti G. A new rapid bedside assay for D-dimer measurement (Simplify D-dimer) in the diagnostic work-up for deep vein thrombosis. J Thromb Haemost. 2003;1(12):2681-3.

16. Subramaniam RM, Heath R, Cox K, Chou T, Stewart J, Sleigh J. Does an immunochromatographic D-dimer exclude acute lower limb deep venous thrombosis? Emerg Med Australas. 2006;18(5-6):457.
17. Neale D, Tovey C, Vali A, Davies S, Myers K, Obiako M, et al. Evaluation of the Simplify d-dimer assay as a screening test for the diagnosis of deep vein thrombosis in an emergency department. Emerg Med J. 2004;21(6):663-6.
18. Emanuel EJ, Fuchs VR. The perfect storm of overutilization. JAMA. 2008;299(23):2789-91.
19. Owens DK, Qaseem A, Chou R, Shekelle P. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. Ann Intern Med. 2011;154(3):174-80.
20. Moons KG. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem. 2010;56(4):537-41.
21. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-9.
22. Schaafsma JD, van der Graaf Y, Rinkel GJE, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. J Clin Epidemiol. 2009;62(12):1248.
23. Geersing GJ, Toll DB, Janssen KJ, Oudega R, Blikman MJ, Wijland R, et al. Diagnostic accuracy and user-friendliness of 5 point-of-care D-dimer tests for the exclusion of deep vein thrombosis. Clin Chem. 2010;56(11):1758-66.
24. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics. 1998;13(4):397-409.
25. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making. 1993;13(4):322-38.
26. Heit JA. Venous stasis syndrome: the long-term burden of deep vein thrombosis. Hosp Med. 2003;64(10):593-8.
27. Kahn SR, Ginsberg JS. Relationship between deep venous thrombosis and the postthrombotic syndrome. Arch Intern Med. 2004;164(1):17-26.
28. Statistics Netherlands. Doodsoorzaken; korte lijst (belangrijke doodsoorzaken), leeftijd, geslacht. [database on the Internet]. Statistics Netherlands. 2011. Available from: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=7052_95&D1=a&D2=a&D3=0&D4=0,41,51,l&HDR=G2,G1,G3&STB=T&VW=T
29. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. University of York, Centre for Health Economics, 1999 November 1999. Report No.
30. van Dongen CJ, Vink R, Hutten BA, Buller HR, Prins MH. Recurrent thromboembolism after treatment with vitamin K antagonists. Arch Intern Med. 2003;163(22):1285-93.
31. Statistics Netherlands. Consumentenprijzen; prijsindex 2006=100. Den Haag/ Heerlen.: Statistics Netherlands; 2011 [cited 2011 Oct]; Available from: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71313ned&LA=NL.
32. Oostenbrink J, Bouwmans C, Koopmanschap M, Rutten F. Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. (in Dutch). Amstelveen2004.
33. Genees- en Hulpmiddelen Informatie Project (GIP databank). Diemen: Health Care

Insurance Board; [cited 2011]; Available from: www.gipdatabank.nl.
34. Hakkaart- van Roijen L, Tan S, Bouwmans C. Handleiding voor kostenonderzoek: methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. Geactualiseerde versie 2010 ed: College voor Zorgverzekeringen.; 2011.
35. Barton GR, Briggs AH, Fenwick EA. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). Value Health. 2008;11(5):886-97.
36. Pink J, Lane S, Pirmohamed M, Hughes DA. Dabigatran etexilate versus warfarin in management of non-valvular atrial fibrillation in UK context: quantitative benefit-harm and economic analyses. BMJ. 2011;343:d6333.
37. Douketis JD, Gu CS, Schulman S, Ghirarduzzi A, Pengo V, Prandoni P. The risk for fatal pulmonary embolism after discontinuing anticoagulant therapy for venous thromboembolism. Ann Intern Med. 2007;147(11):766-74.
38. Murin S, Romano PS, White RH. Comparison of outcomes after hospitalization for deep venous thrombosis or pulmonary embolism. Thromb Haemost. 2002;88(3):407-14.
39. Douketis JD, Kearon C, Bates S, Duku EK, Ginsberg JS. Risk of fatal pulmonary embolism in patients with treated venous thromboembolism. JAMA. 1998;279(6):458-62.
40. Prandoni P, Noventa F, Ghirarduzzi A, Pengo V, Bernardi E, Pesavento R, et al. The risk of recurrent venous thromboembolism after discontinuing anticoagulation in patients with acute proximal deep vein thrombosis or pulmonary embolism. A prospective cohort study in 1,626 patients. Haematologica. 2007;92(2):199-205.
41. Prandoni P, Lensing AW, Cogo A, Cuppini S, Villalta S, Carta M, et al. The long-term clinical course of acute deep venous thrombosis. Ann Intern Med. 1996;125(1):1-7.
42. Nieto JA, Solano R, Ruiz-Ribo MD, Ruiz-Gimenez N, Prandoni P, Kearon C, et al. Fatal bleeding in patients receiving anticoagulant therapy for venous thromboembolism: findings from the RIETE registry. J Thromb Haemost. 2010;8(6):1216-22.
43. Gibson NS, Schellong SM, Kheir DY, Beyer-Westendorf J, Gallus AS, McRae S, et al. Safety and sensitivity of two ultrasound strategies in patients with clinically suspected deep venous thrombosis: a prospective management study. J Thromb Haemost. 2009;7(12):2035-41.
44. Caprini JA, Botteman MF, Stephens JM, Nadipelli V, Ewing MM, Brandt S, et al. Economic burden of long-term complications of deep vein thrombosis after total hip replacement surgery in the United States. Value Health. 2003;6(1):59-74.
45. Bergman L, van der Meulen JH, Limburg M, Habbema JD. Costs of medical care after first-ever stroke in The Netherlands. Stroke. 1995;26(10):1830-6.

APPENDIX 1 FLOW CHART OF SYSTEMATIC SEARCH IN PUBMED DATABASE



POCT= point-of-care test

APPENDIX 2 DATA EXTRACTION AND 2X2 TABLES

Study	Patients (n)	Mean age (years, (SD or range))	Health care setting	Clinical Decision Rule used	DVT Prevalence (%)	Cut-off value used
Triage						
Geersing et al., 2010	577	58 (16)	prim	None	12.3	350 ug/L FEU
Baker et al., 2010	101	62 (17-92)	sec	Wells	10.9	400 ug/L FEU
Cardiac						
Legnani et al., 2003	80	n.r. (21-94)	sec	Wells	40.0	500 ug/L[3]
Bucek et al., 2001	85	59 (17)	sec	Wells	41.2	500 ug/L[3]
Geersing et al., 2010	577	58 (16)	prim	None	12.3	500 ug/L FEU
Dempfle et al., 2006	560	58 (17)	sec	Wells	39.8	500 ug/L FEU
Nycocard						
Gardiner et al., 2005	397	n.r.	sec	Wells	20.2	300 ug/L[3]
Larsen et al., 2002	113	57 (20-99)	sec	None	43.4	300 ug/L[3]
Simplify						
Büller et al., 2009	997	58 (17)	prim	Oudega	13.4	80 ug/L[3]
Cini et al., 2003	120	64 (20-96)	sec	Wells	29.2	80 ug/L[3]
Geersing et al., 2010	577	58 (16)	prim	None	12.3	80 ug/L[3]
Subramaniam et al., 2006	453	56 (20)	sec	Hamilton	19.2	80 ug/L[3]
Neale et al., 2004	187	n.r.	sec	None	27.3	80 ug/L[3]

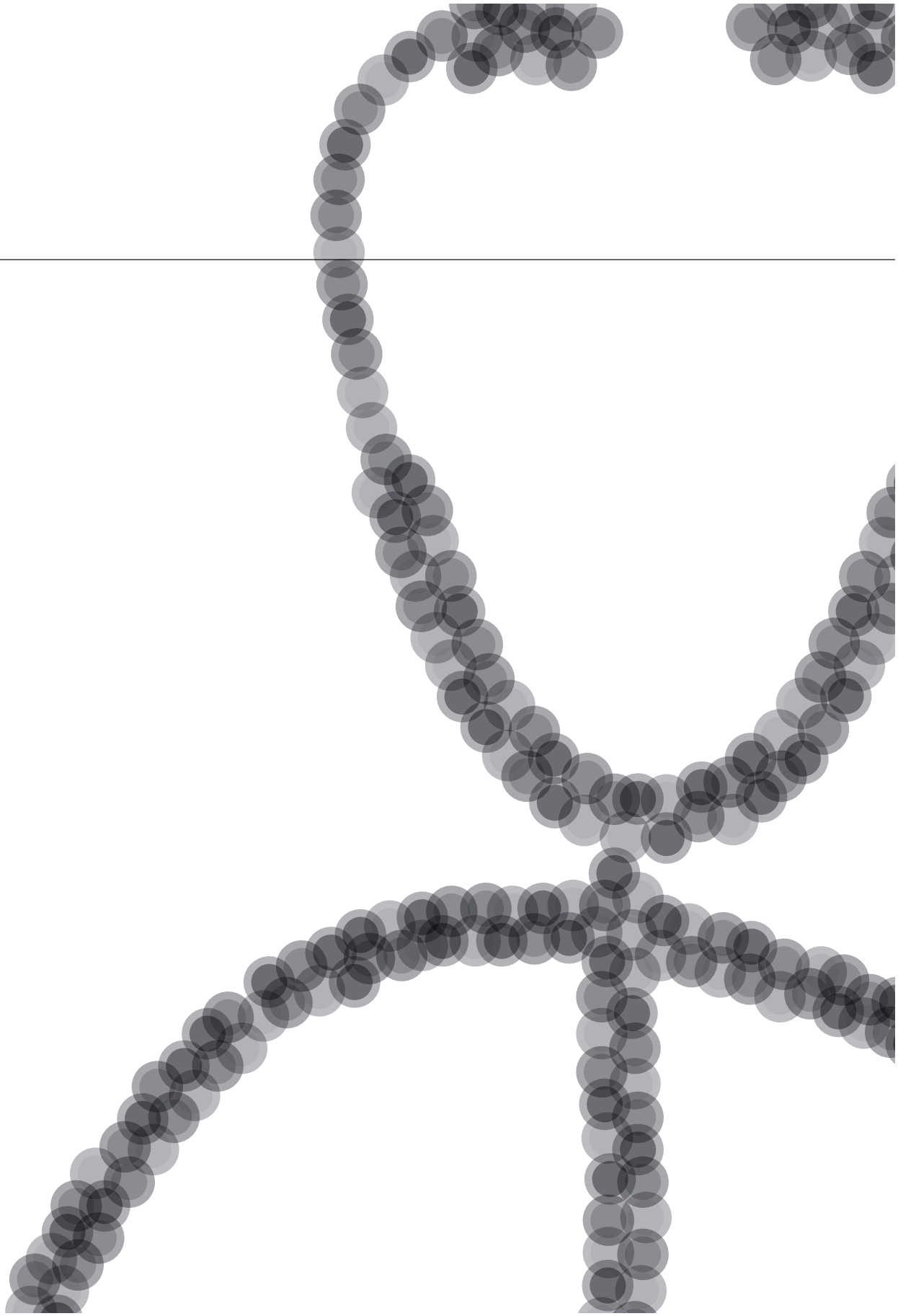
[1] Follow-up period of 3 months. [2] 2 by 2 table is extracted from the original report of the study instead of the paper. [3] = D-dimer measurement unit not reported.

n.r. = not reported; prim= primary care; sec = secondary care; FEU= Fibrinogen Equivalent Units; NPV= negative predictive value; PPV= positive predictive value; TP= true positive; FP= false positive; FN= false negative; TN= true negative; 95% CI= 95 % confidence interval

Reference standard 1=CUS 2=venography 3=follow-up period[1]	Sensitivity (95% CI)	Specificity (95% CI)	NPV	PPV	TP	FP	FN	TN
1	0.972 (0.922 to 0.996)	0.480 (0.438 to 0.523)	0.99	0.21	69	2	263	243
1	1.00 (n.a.)	0.321 (0.233 to 0.420)	1.00	0.15	11	0	61	29
1+3	0.969 (0.896 to 0.999)	0.605 (0.460 to 0.737)	0.97	0.62	31	1	19	29
1+2	0.885 (0.772 to 0.968)	0.500 (0.365 to 0.639)	0.86	0.55	31	4	25	25
1	0.943 (0.883 to 0.983)	0.620 (0.576 to 0.661)	0.99	0.26	67	4	192	314
1	0.968 (0.942 to 0.988)	0.610 (0.557 to 0.664)	0.97	0.62	216	7	132	205
1+2	0.900 (0.824 to 0.953	0.448 (0.394 to 0.503)	0.95	0.29	72	8	175	142
1	0.633 (0.494 to 0.760)	0.674 (0.557 to 0.781)	0.70	0.60	31	18	21	43
1+3	0.859 (0.797 to 0.914)	0.617 (0.585 to 0.651)	0.97	0.26	115	19	331	532
1+3	1.00 (n.a.)	0.532 (0.428 to 0.633)	1.00	0.47	35	0	40	45
1	0.914 (0.833 to 0.967)	0.641 (0.599 to 0.683)	0.98	0.26	65	6	182	324
1+3	0.851 (0.772 to 0.917)	0.585 (0.534 to 0.639)	0.94	0.33	74	13	152	214
2	0.940 (0.853 to 0.987)	0.404 (0.325 to 0.485)	0.95	0.37	48	3	81	55

APPENDIX 3 POINT-OF-CARE D-DIMER TEST CHARACTERISTICS

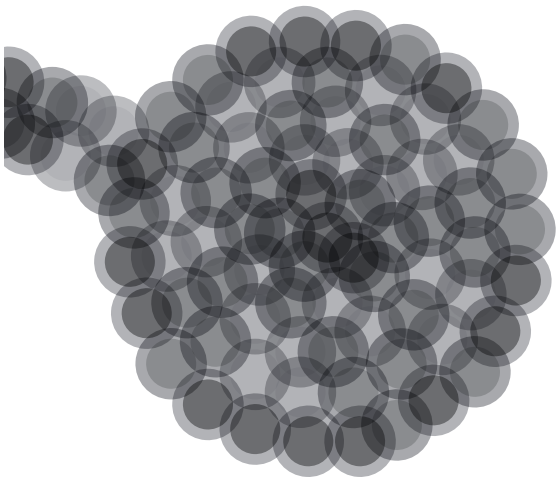
"Point-of-care" D-dimer assays	
Triage	Triage D-dimer (Biosite, San Diego, CA, USA) is performed with 250ul of EDTA whole blood, making use of a small portable device, the Triage Meter Plus. Within 15 minutes, a quantitative result is displayed within a range of 100-5,000 ug/l D-DU, with a cut-off value of 350 ug/l D-DU.
Cardiac	Cardiac D-dimer (Roche Diagnostics, Mannheim, Germany) requires 150 ul whole venous blood, anticoagulated with Li-heparin, and the Cobas h232, a small portable instrument. Within a range of 100-4000 ug/l FEU, the analyzer displays a quantitative value in less than 15 minutes. The manufacturer's recommended cut-off value is 500 ug/l FEU.
Nycocard	Nycocard D-dimer (Axis Shield PoC AS, Oslo, Norway), is performed with 50 ul of plasma and quantitatively displayed on the NycoCard Reader II with a range of 100-20,000 ug/L FEU within 3 minutes. A threshold of 300 ug/l FEU is recommended by its manufacturer.
Simplify	Clearview Simplify D-dimer (Inverness Medical, Bedford, UK) is a qualitative assay and requires 35 ul of capillary whole blood. When it is mixed with two drops of test reagent, a purple pink-stained line is visible in case of a D-dimer level above 80 ug/l within 10 minutes (assumed to be a positive test result). A test is only valid if a second control line is visible as well.



CHAPTER 6

A DIFFERENT VIEW ON RISK FACTORS FOR RECURRENT VENOUS THROMBOEMBOLISM

Janneke MT Hendriksen, Geert-Jan Geersing, Roger EG Schutgens



Manuscript in preparation

PATIENT VIGNETTE

A 52 year-old man suffered from an unprovoked pulmonary embolism six months ago. After one week of combined use of low molecular weight heparin (LMWH) and acenocoumarol (vitamin K antagonist; VKA), single anticoagulant treatment was continued by means of acenocoumarol with an INR (International Normalized Ratio) targeted between 2.5-3.5. During the outpatient visit six months post-event, the anticoagulant therapy duration is discussed. The patient has ambivalent feelings towards treatment withdrawal after the standard six months of treatment: use of medication, the concomitant bleeding risk and the regular INR controls are reasons to consider discontinuation of treatment. However, he has also read about the significant proportion of patients with unprovoked venous thrombosis that develop a recurrent event after VKA withdrawal. And he definitely does not want to go through such a frightening experience again. His treating physician suggests performing a risk assessment.

INTRODUCTION

There is an ongoing discussion on the optimal duration of anticoagulation therapy after experiencing a venous thromboembolic event (VTE). Treatment is initially initiated to prevent thrombus extension post-event aiming to yield symptom relief and even prevention of acute mortality. Nevertheless, ongoing treatment has a different goal, namely reducing recurrent events in future. These benefits should be balanced against the potential complications of treatment. The 2009 Dutch guideline (CBO consensus) states that the treatment duration is (at least partially) dependent on the presence of transient risk factors in relation to the thrombotic event. (1) For a first provoked VTE, the suggested treatment duration is 3 months, given the fact that the risk of VTE recurrence is low after discontinuation (2). In case of unprovoked thrombosis, six months of anticoagulant treatment is recommended, given the fact that the recurrence risk is higher. (2) However, because of this higher recurrence risk, American guidelines are more stringent in their recommendations and recommend considering a prolonged duration of treatment beyond 3-6 months in a larger group of patients. (3) Currently, the Dutch guidelines are under revision, guided by the present evidence on the optimal treatment duration.

In this paper, we discuss the view on the optimal duration of anticoagulant therapy in VTE and the possibilities to individually tailor this optimal treatment duration. We anticipate a paradigm shift in the treatment of VTE, analogue to the evolving guidelines for anticoagulant use in atrial fibrillation. The focus is expected to shift towards the definition and subsequent identification of subsets of patients with predicted low recurrence risk whom will benefit sufficiently from a short treatment duration, versus an indefinite duration of treatment in all other VTE patients. Furthermore, we consider the recently introduced prediction models as potential tools to truly estimate and interpret the value of individual risk factors for recurrent VTE, in order to improve individual risk prediction.

EPIDEMIOLOGY

Venous thromboembolic disease is listed amongst the conditions that account for most cardiovascular deaths each year, ranked third next to stroke and myocardial infarction. It is estimated that around 500,000 persons die each year as a consequence of venous thromboembolism in Europe. VTE incidence is 1.92 per 1000 person years. (4)

The cornerstone of initial VTE treatment consists of anticoagulant therapy for the course of several months. The aim is twofold: first, prevention of further thrombus growth, and subsequent relief of symptoms and reduction of mortality, and second the prevention of recurrent events. The incidence of recurrent VTE events after anticoagulant withdrawal is 7.6 per 100 patient years with a case-fatality rate of 3.6%. (6) A recent study showed that prolonged anticoagulant treatment (at least longer than three months) was associated with decrease in all-cause mortality (HR 0.82 (95% CI 0.73-0.92) if prolonged treatment >12 months). (7)

TABLE 1 OVERVIEW OF RANDOMIZED STUDIES ON THE EFFECT OF PROLONGED ANTICOAGULANT TREATMENT IN A FIRST (UN)PROVOKED VENOUS THROMBOEMBOLIC EVENT.

	RCT	Domain	N=
Initial duration treatment of at least 3 months			
Schulman, 1995 (29)	VKA discontinuation at 6 weeks or 6 months	1st (un)provoked VTE	897
Agnelli, 2001 (30)	Withdrawal at 3 months VKA vs. 9 months prolonged treatment	1st unprovoked proximal DVT, after 3 months of VKA	267
Pinede, 2001 (31)	3 months vs. 6 months VKA	1st distal or proximal DVT and/or PE	736
Agnelli, 2003 (32)	Intervention: prolonged treatment up to 6 months (provoked) or 12 months (unprovoked)	1st provoked or unprovoked PE, after 3 months of VKA	326
Campbell 2007 (33)	3 vs. 6 months VKA	DVT and/or PE	749
Prolonged anticoagulant treatment after 3-6 months			
Kearon 1999 (34)	Prolonged treatment VKA vs. placebo for 24 months	1st unprovoked VTE, after 3 months VKA	162
Ridker, 2003 (11)	Prolonged treatment with low-intensity VKA (INR 1.5-2.0) vs. placebo	Unprovoked VTE, mean 6,5 months VKA treatment completed	508
Farraj, 2004 (35)	24 months prolonged VKA treatment vs. VKA stopping after 6 months	1st episode VTE	64
Palareti, 2006 (36)	VKA discontinuation if low D-dimer. If elevated D-dimer randomization: discontinuation or prolonged VKA treatment	1st unprovoked proximal DVT or PE, 3 months VKA-treatment completed	608, 223 (36.7%) D-dimer high
Siragusa, 2008 (37)	prolonged 9 months VKA if residual thrombosis (RVT) vs. no prolonged VKA	1st DVT, 3 months VKA treatment completed	258
Couturaud, 2015 (38)	Prolonged treatment VKA vs. placebo for 18 months	1st unprovoked PE, after 6 months VKA	371
Aspirin			
Becattini, 2012 (39)	Aspirin 100mg daily vs. placebo.	1st unprovoked VTE, 6-18 months VKA used	402
Brighton, 2012 (40)	aspirine vs. placebo	1st unprovoked VTE, 6-18 months VKA used	822

RCT= randomized controlled trial; VTE= venous thromboembolism; VKA= vitamin K antagonist; vs. = versus; DVT= deep vein thrombosis; PE= pulmonary embolism, HR= hazard ratio; RR= relative risk; INR= international normalized ratio; mg= milligram

	Recurrent VTE Controls	Recurrent VTE Intervention	Outcome measure	Major bleeding Controls	Major bleeding Intervention
	80/443 (18.1%)	43/ 454 (9.5%)	6-wk OR 2.1 (1.4 to 3.1)	1/443 (0.2%)	5/454 (1.1%)
	21/133 (15.8%)	21/134 (15.7%)	HR 0.99 (0.57 to 1.73)	11/133 (8.3%)	1/134 (0.7%)
	23/375 (6.4%)	26/ 361 (7.4%)	RR 0.87 (0.50 to 1.49)	6/ 375 (1.7%)	10/361 (2.8%)
	18/161 (11.2%) unprovoked PE: 11/91 (12.2%)	15/165 (9.1%) unprovoked PE: 11/90 (12.2%)	RR 0.81 (0.42 to 1.56)	0/161 (0.0%)	3/165 (1.8%)
	31/369 (8%)	29/380 (8%)	p=0.80 for difference. (-3.1% to 4.7%)	0/369 (0.0%)	8/380 (2.1%)
	17/83 (27.4%/ patient-year)	1/79 (1.3%/ patient-year)		0/83 (0.0%)	3/79 (3.8%)
	37/253 (7.2/100 patient-years)	14/255 (2.6/100 patient-years)	HR 0.36 (0.19 to 0.67)	2/253 (0.4/100 patient-years)	5/255 (0.9/100 patient-years)
	7/32 (21%)	1/32 (3%)		2/32 (6%) (in 1st 3 months after start VKA)	2/32 (6%) (in 1st 3 months after start VKA)
	223 (36.7%) abnormal D-dimer. Stopped VKA: 18/120 (15.0%) low D-dimer: 24/385 (6.2%)	3/103 (2.9%)	HR 2.27 (1.15 to 4.46) HR 4.26 (1.23 to 14.6)	0/505 (0.0%)	1/103 (1.0%)
	25/92 (27.2%)	17/88 (19.3%)	HR 1.58 (0.85 to 2.93)	1/92 (1.1%)	2/88 (2.3%)
	25/187 (13.5%	3/184 (1.6%)	HR 0.15 (0.05 to 0.43)	1/187 (0.5%)	4/184 (2.2%)
	43/197 (11.2%/ patient-year)	28/205 (6.6%/ patient-year)	HR 0.58 (0.36 to 0.93)	1/197 (0.5%)	1/205 (0.5%)
	73/411 (6.5%/ patient-year)	57/ 411 (4.8%/ patient-year)	HR 0.74 (0.52 to 1.05)	0.6% / year	1.1% / year

Experiencing a recurrent VTE event is associated with multiple adverse outcomes, including a considerable mortality risk. For example, the risk of developing a post-thrombotic syndrome increases from 6.8% (1st DVT) up to 18.1% (recurrent event) in the first year after the DVT event. Furthermore, the risk of developing venous stasis, ulcers and chronic thromboembolic pulmonary hypertension doubles. (8) Recurrent pulmonary embolism has a significant impact on daily functioning of the patient, up to three years post-event. (9) In a Dutch study, patients with recurrent PE performed worse on social functioning, their emotional and physical role, their general health state and activity levels in comparison with the general population. (10) Total health care costs that are associated with recurrent VTE events are estimated to be 33.617 euro per patient per year. (8)

RECURRENT THROMBOSIS: URGE FOR PREVENTION

Thus, in the light of the potential detrimental consequences for patients, the importance of recurrent VTE prevention is evident. Roughly, two methods to prevent recurrent VTE are available. Firstly, by optimal non-pharmacological prevention (e.g. discontinuation of hormonal substitution therapy, quit smoking, weight reduction) VTE recurrences can be prevented. And secondly, by optimization of the anticoagulant treatment in terms of duration and intensity, the formation and extension of thrombi can be restrained.

Multiple studies have shown that the risk of recurrent VTE decreases while on (prolonged) anticoagulant treatment. These studies have been described in greater detail in *Table 1*. Prolonged use of vitamin K antagonists (VKA) reduces the risk of recurrent VTE with 90% compared to placebo. (11) Comparable results were found in the recent studies with the direct oral anticoagulants dabigatran, rivaroxaban and apixaban. (12) A disadvantage of trials on the prolonged duration of VKA is the fact that the participants were randomization in two groups (continued or discontinued anticoagulant treatment) without taking into account the individual recurrence risk. As such, it is expected that individual patients receive either too long or too short anticoagulant treatment. Or, in other words, the positive effect on the reduction of the number of recurrent cases in these studies is mainly driven by the reduction of VTE in patients with the highest recurrence risk, while patients with a low predicted recurrence risk do not benefit from prolonged treatment. In fact, these low-risk patients are exposed to potentially detrimental side-effects of anticoagulant treatment and the subsequent risk of bleeding complications. As a consequence, the focus has been shifted to the prediction of individual recurrence risk based on several risk factors recently.

KNOWN RISK FACTORS OF RECURRENT VENOUS THROMBOSIS

The list of known risk factors for recurrent VTE has been expanded to a list including over 50 factors in the course of the years. (13) These risk factors can be subdivided into transient and permanent risk factors. Transient risk factors are for example bed rest, long haul flights, immobilisation, surgery, presence of a venous

catheter, obesity, pregnancy and use of medication like specific antipsychotics, oral contraceptives and hormonal therapy. Permanent risk factors include age, malignancy and thrombophilia. For each individual risk factor, multiple (etiological) studies can be found, some with results that are comparable, but often with conflicting results regarding the risk on thrombosis and recurrence. Moreover, some factors play an important role in both first and recurrent events, for example the presence of malignancy, whereas other factors are only of influence for developing a first thromboembolic event (e.g. factor V Leiden), or a recurrent event (for example residual thrombosis). And furthermore, international experts disagree on the clinical relevance of certain risk factors regularly. For instance, a patient with an unprovoked thrombotic event with a protein C deficiency should be treated for an indefinite duration according to one physician (13) , whereas another treating physician would say that evidence is too limited to initiate anticoagulant therapy for this prolonged duration in a patient with a thrombotic event related to hereditary thrombophilia. (12) Due to the conflicting evidence and the difficulties to value individual risk factors, it can sometimes be difficult for treating physicians to decide what to do for a specific patient.

CLASSIFICATION OF RISK FACTORS FOR RECURRENT VTE

To provide tools for clinical practice, we describe another classification of the risk factors for recurrent VTE. It has to be stressed that this classification does not cover highly specific patient groups, like patients with myeloproliferative disorders or the antiphospholipid syndrome.

1. Evident risk factors

This group includes non-modifiable risk factors with none or only limited discussion on their clinical impact.

Distal versus proximal

The location of the initial VTE event is predictive of the risk of recurrence. A distal DVT (calf veins) is associated with a relatively low recurrence rate; compared to a proximal VTE, the recurrence risk of a distal DVT is 2-5 times lower. (14, 15)

Malignancy

When a thrombotic event occurs during the active phase of a malignant disease (latest treatment ‹ 6 months ago), the risk of recurrence is very high. A recent study reports a VTE recurrence risk of 52% within 10 years in patients with active cancer. (16)

Previous venous thromboembolic event

The 5-year cumulative risk of recurrent VTE was 21.6% after a first VTE and 27.9% after a second unprovoked VTE. (17) In the DURAC study, patients with a second unprovoked VTE were randomized between a treatment period of 6 months or

treatment for undetermined duration. The risk of recurrence dropped from 20.7% to 2.6% in favour of prolonged treatment, with a trade-off of a non-significant increase of bleeding events (from 2.7% to 8.6%). (18) Despite a reduction of total mortality from 14.6% to 8.7%, this reduction did not meet statistical significance either.

2. Modifiable risk factors

This is a highly appealing group to physicians in daily practice, since the potential modifiable nature of the risk factors offers the physician interventions in the consultation room. If a risk factor can be identified as cause of the thrombotic event, or as a potential risk for recurrence, the elimination of this risk factor can lead to significant risk reduction of recurrence. Obviously, these potentially modifiable risk factors are almost always transient. Two examples are given:

Oestrogens

The use of hormonal substitution therapy, including oral contraceptives, should be considered as a transient and modifiable risk. A recent study explored these factors in 630 women after a first VTE event and demonstrated the temporal character of these risk factors. Of these 630 women, 333 used a form of oestrogen therapy at the moment of the initial VTE event; the oestrogen therapy was discontinued in all. Anticoagulant treatment duration varied between 3 and 18 months (mean 7 (3) months). The recurrence risk after 5 years was 17% in the 297 women who had never taken any oestrogens versus 7% in the 333 women who stopped oestrogen therapy after the VTE event. (19)

Long-haul flights

Despite the fact that the absolute risk of thrombosis is low during long-haul flights, this risk increases gradually with increased duration of the flight. In comparison with a flight duration of 6 hours, a travel duration of 8 hours is associated with a VTE risk that is 2.3 times as high. This risk is even 3.6 times higher in patients with another risk factor for recurrence, like a previous thrombotic event. (20)

3. Other risk factors

Provoked versus unprovoked

In case the initial VTE cannot be classified as provoked by a modifiable or evident risk factor, the event is called "unprovoked". In these patients, the risk of recurrence is higher than compared to the risk in patients with a VTE caused by a modifiable risk factor, provided that the provoking factor is not present anymore. (14): To illustrate this, the VTE recurrence risk is only 0.7% per patient year if the index event was provoked by a surgical procedure; this risk is 4.2% per patient year in case of another, non-surgical, provoking factor (pregnancy, (oral) contraceptives, long-haul travel) and 7.4% per patient year in the absence of any of these provoking factors. (2, 12). It should be noted that "unprovoked" does not imply that an unprovoked event is entirely "spontaneously" in origin. Instead, the

term unprovoked should be regarded as a term to describe more uncertainty in the recurrence risk compared to the risks for the evident and modifiable risk factors.

Residual thrombosis

In 30%-50% of all DVT patients, residual thrombosis is present on ultrasonography of the affected vein after 1 year. (21) The value of this radiological finding as a potential risk factor for recurrent VTE is not entirely clear. In a meta-analysis of 14 studies, the presence of residual thrombosis at the moment of VKA withdrawal appeared to be a minor risk factor for recurrent VTE in patients with all types of DVT, including provoked and unprovoked events (OR 1.5 (95% CI 1.1-2.0)). However, the presence of residual thrombosis in the specific patient group with unprovoked DVT does not seem to be predictive for VTE recurrence (OR 1.24 (95% CI 0.9-1.7). (22)

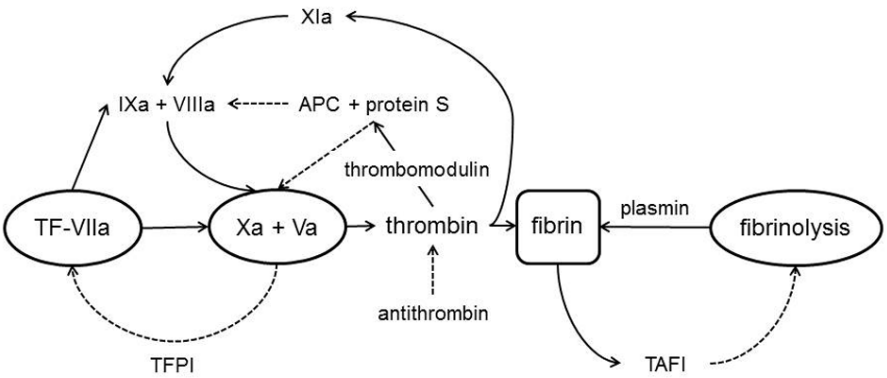
Increased coagulation system activity

Numerous permanent risk factors have their origin in their role in the secondary coagulation cascade. Examples include elevated concentrations of coagulation factors VIII, XI and the von Willebrand factor, mutations in the prothrombin gen (FII mutation) and mutations in FV (FVLeiden) causing a decreased functioning of the inhibition of the protein C and S systems. Furthermore, deficiencies in endogenous anticoagulants like protein C, protein S and antithrombin lead to an increased thrombotic tendency. As all these factors have their repercussions on the final product of the coagulation cascade, i.e. fibrin formation, it is attractive to use a general screenings test of fibrin formation to gain insight in the coagulation activity. D-dimer assays quantify this fibrin degradation product levels. As such, D-dimer is an indirect measure of fibrin formation (see *Figure 1*) and consequently the D-dimer can be used to predict recurrent VTE as well. The result of a D-dimer test performed one month after VKA withdrawal is predictive for the risk of VTE recurrence: a patient with an elevated D-dimer level (>500 ng/ml) has a 2.5 times higher risk of recurrence compared to a patient with a non-elevated D-dimer level. (23) This has been investigated in the DULCIS study. Anticoagulant treatment was stopped in patients with a normal D-dimer level (in 52% of cases) after 6 months of treatment; in patients with an elevated D-dimer level any time during the three months following treatment cessation, the anticoagulant therapy was resumed. (24) VTE recurrence risk was 3.0% per 100 patient years in the group that stopped after 6 months of treatment, versus 0.7% per 100 patient years in the group with a prolonged anticoagulant treatment duration. Some patients refused to resume the anticoagulant treatment despite an elevated D-dimer level (11% of the total number of participants in the study). Recurrence risk was 8.8% per 100 patient years in this group.

4. Prediction models

It is evident that it is not always possible to identify a univocal risk factor as predictor of recurrence. Therefore, the ultimate aim is to combine certain risk

FIGURE 1 SECONDARY COAGULATION CASCADE



TF= tissue factor; APC= activated protein C; TFPI= tissue factor pathway inhibitor; TAFI= thrombin activatable fibrinolysis inhibitor

factors to improve the predictive potential. To do this, prediction models can be used. Currently, three prediction models have been developed that predict VTE recurrence risk (see *Table 2*). All models provide an estimated absolute VTE recurrence risk per patient, based on the combination of a certain set of patient characteristics, including the biomarker D-dimer. The definition of the threshold of an increased recurrence risk is not unequivocal. Internationally, a 1-years recurrence risk of 5% is suggested to be an acceptable threshold. (25) This percentage is based on the fact that a 1-year recurrence risk ‹5% is accepted in patients with a provoked VTE after discontinuation of anticoagulation therapy, but that the 1-year recurrence rate of 10% in unselected VTE patients after treatment cessation is considered to be too high. (25)

THE PARADIGM SHIFT: TOWARDS IDENTIFICATION OF LOW RISK (INSTEAD OF HIGH-RISK) PATIENTS

Three recent developments have led to an altered view on the treatment of venous thromboembolism. Firstly, the introduction of the direct oral anticoagulants (DOACs). DOACs have entered the VTE treatment spectrum with comparable effectiveness as compared to the conventional vitamin K antagonists. (26) A recent meta-analysis demonstrates that the risk of clinically major bleeding is lower for the use of DOACs compared to VKA: 1.8% versus 3.1% in favour of the DOACs. Also the case fatality of these bleeding events was lower (6.1% versus 10.4%). (27) Considering these

TABLE 2 OVERVIEW OF THE VTE RECURRENCE PROGNOSTIC PREDICTION MODELS CURRENTLY DEVELOPED

	Rodger et al (41)	Vienna Prediction model (42)	DASH-score (43)
Predictors	- Post-thrombotic syndrome - D-dimer ≥ 250 µg/L # - BMI ≥ 30 kg/m2 - Age ≥ 65 years	- Gender - D-dimer (µg/L) - Location index VTE	- Gender - Abnormal$ D-dimer - Hormonal (replacement) therapy - Age ≤ 50 years
Annual VTE recurrence risk	Low risk: score <2: <3% High risk: score ≥2: 10.2%	Low risk: <5% High risk: ≥5%	Low risk: score ≤1: 3.1% High risk: score =2: 6.4% score >2: 12.3%
Notes	only in women (all men continue treatment)	nomogram & online calculator available	

BMI= Body Mass Index; µg/L = micrograms per litre; kg/m2 = kilograms per square meter.
# D-dimer measured while taking vitamin K antagonist. $ Abnormal D-dimer= positive D-dimer test result if qualitative test used, or ≥500 µg/L if quantitative test used, 3-5 weeks after discontinuation anticoagulant treatment.

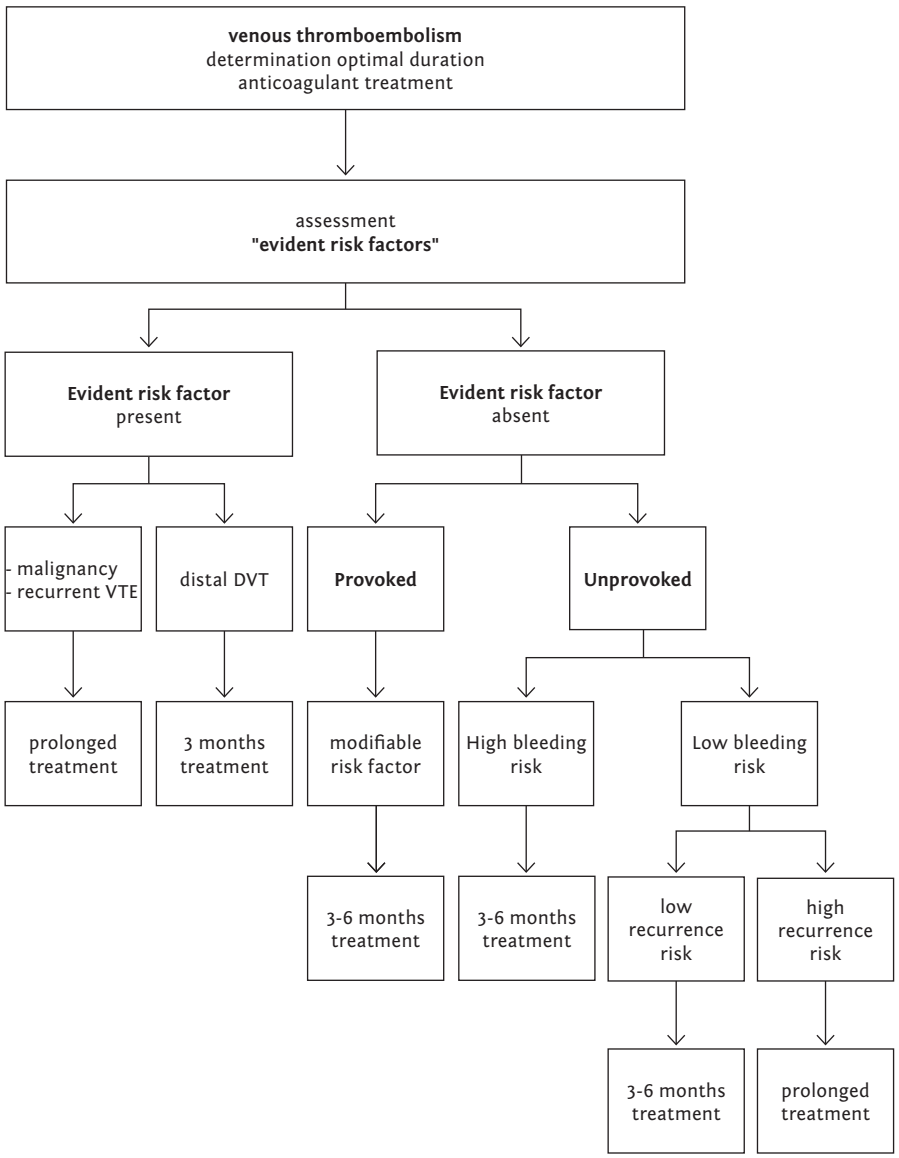
observations, the expected advantages of the DOACs over VKAs might tip the scale towards prolonged anticoagulant treatment duration in a subset of VTE patients.

Secondly, the fact that we are increasingly capable to predict risks on an individual level (see above) and thirdly, the growing understanding that venous thromboembolism is, in a substantial proportion of patients, a manifestation of an underlying (chronic) thrombosis tendency. (5) The chance of developing a new thrombosis will persistently be present. As such, prolonged treatment should be considered, at least in a (substantial) subgroup of VTE patients.

Whereas a "one size fits all" treatment duration of 3 to 6 months has been the norm until now, we propose a shift towards a risk tailored approach, in which the focus is on the identification of low-risk patients for recurrences and high-risk patients for bleeding. The former patients can safely stop anticoagulant treatment after the standard treatment duration. Consequently, in all patients not identified as low-risk, a prolonged treatment should be considered.

A comparable paradigm shift has been taken place several years ago in the prevention of ischemic stroke in atrial fibrillation. The introduction of the CHA2DS2-VASc score has led to the identification of patients that do not benefit from anticoagulant treatment, that is, the subgroup of patients that is at truly low risk of ischemic stroke.

FIGURE 2 PROPOSED FLOW CHART OF THE PROGNOSTIC RISK STRATIFICATION OF RECURRENT VTE AND SUBSEQUENT DETERMINATION OF THE OPTIMAL DURATION OF ANTICOAGULANT TREATMENT



VTE= venous thromboembolism; DVT= deep venous thrombosis

We expect a similar trend in patients with a first VTE. In these patients, first the presence of evident risk factors is assessed. If one of these factors is present, there is no discussion on the optimal treatment duration. If not, the presence of modifiable risk factors is assessed. In patients with modifiable risk factors, a treatment duration of three months will be sufficient on the condition that these modifiable factors like contraceptive use are not present anymore. These factors can also be used in patient counselling, like long-haul flight related prophylaxis, smoking cessation and weight reduction. In the remaining patients, primarily those with a first unprovoked VTE event, a risk assessment will take place after three to six months, based on one of the three existing prediction models. Anticoagulant treatment will be stopped after 3-6 months in patients with a low predicted recurrence risk, or with a high predicted bleeding risk. The threshold for a high predicted recurrence risk is consensus-based set at 5% per year. (25) The use of anticoagulants will be prolonged for an undefined period in all other patients. Prolonged duration of treatment requires explicit agreements concerning patient monitoring and the coordination physician. One can think of a yearly evaluation of the benefits and disadvantages of continued anticoagulant treatment. A flow chart of the risk stratification is presented in *Figure 2*.

Obviously, the bleeding risk in each individual patient should be definitely taken into account when deciding upon prolonged treatment duration based on the outcomes of a prediction model. Several prediction models, like the HAS-BLED score (28), have been developed, predominantly in patients with atrial fibrillation. Nevertheless, the application of these scores in patients using VKA because of an indication other than atrial fibrillation (such as VTE), or in a Dutch primary care setting, has not been investigated yet.

Finally, an important aspect for physicians treating patients with VTE is the so-called "shared decision making". Ultimately, it is the patients' choice whether or not to stop or prolong anticoagulation therapy. Therefore, it is of utmost importance that patients know about the risks and benefits and are able to discuss this freely with their physician.

The concept of a risk-tailored anticoagulant treatment is currently evaluated in a randomized controlled trial (VISTA study, Dutch Trial Register number 2680). In this study, patients with an unprovoked VTE are randomized to standard care conform current guidelines or a risk-tailored treatment based on the outcomes of the Vienna Prediction Model. In case of an elevated predicted risk (>5% per year) the patient and treating physician are advised to prolong the initial treatment duration with two more years. Study inclusion is expected to be finished by the end of 2015, final results are expected to be presented after completion of follow-up in the course of 2017.

CONCLUSIONS

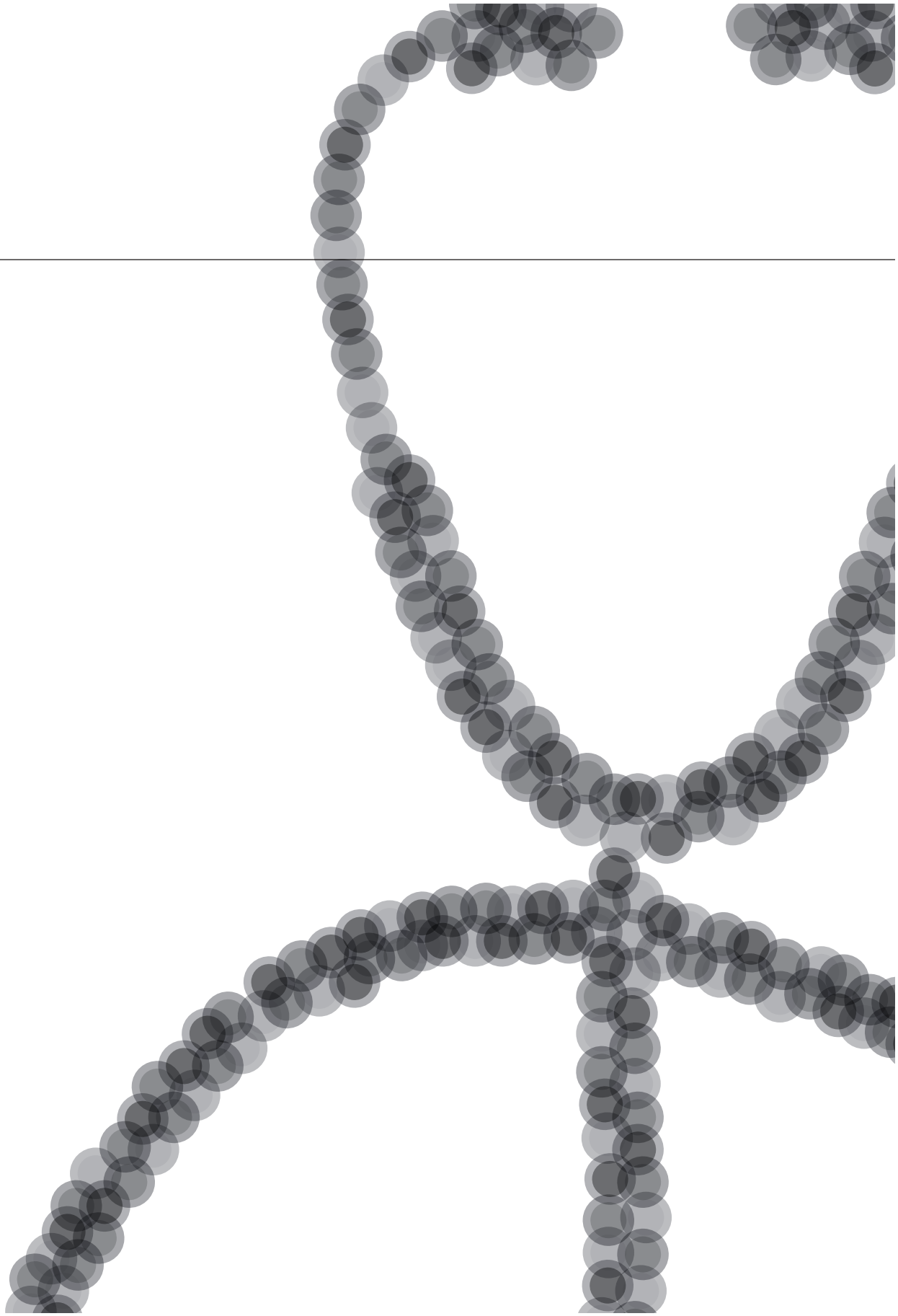
Our understandings in risk factors for recurrent venous thromboembolism and the concordant consequences for the optimal treatment duration are in motion. The use of prediction models is expected to increase, with the aim to perform accurate individual risk assessment and tailor anticoagulant treatment duration accordingly after VKA withdrawal after an unprovoked venous thromboembolic event. Prolonged treatment duration is expected to be prescribed more often, with the aim to decrease the burden of recurrent venous thromboembolism.

REFERENCES

1. Kwaliteitsinstituut voor de Gezondheidszorg CBO. Richtlijn: Diagnostiek, preventie en behandeling van veneuze trombo-embolie en secundaire preventie arteriele trombose. Utrecht 2008.
2. Iorio A, Kearon C, Filippucci E, Marcucci M, Macura A, Pengo V, et al. Risk of recurrence after a first episode of symptomatic venous thromboembolism provoked by a transient risk factor: a systematic review. Arch Intern Med. 2010;170(19):1710-6.
3. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-based clinical practice guidelines. Chest. 2012;141(2_suppl):e419S-e94S.
4. Cushman M, Tsai AW, White RH, Heckbert SR, Rosamond WD, Enright P, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. Am J Med. 2004;117(1):19-25.
5. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015;advance online publication.
6. Carrier M, Le Gal Gg, Wells PS, Rodger MA. Systematic review: case-fatality rates of recurrent venous thromboembolism and major bleeding events among patients treated for venous thromboembolism. Ann Intern Med. 2010;152(9):578-89.
7. Klil-Drori A, Coulombe J, Suissa S, Tagalakis V. Extended durations of anticoagulation and mortality following incident venous thromboembolism in the QVTE cohort. J Thromb Haemost. 2015;13(suppl S2):243.
8. Lefebvre P, Laliberté F, Nutescu EA, Duh MS, LaMori J, Bookhart BK, et al. All-cause and disease-related health care costs associated with recurrent venous thromboembolism. Thromb Haemost. 2013;110(12):1288-97.
9. Klok FA, Cohn DM, Middeldorp S, Scharloo M, Buller HR, van Kralingen KW, et al. Quality of life after pulmonary embolism: validation of the PEmb-QoL Questionnaire. J Thromb Haemost. 2010;8(3):523-32.
10. van Es J, den Exter PL, Kaptein AA, Andela CD, Erkens PMG, Klok FA, et al. Quality of life after pulmonary embolism as assessed with SF-36 and PEmb-QoL. Thromb Res. 2013;132(5):500-5.
11. Ridker PM, Goldhaber SZ, Danielson E, Rosenberg Y, Eby CS, Deitcher SR, et al. Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. N Engl J Med. 2003;348(15):1425-34.
12. Kearon C, Akl EA. Duration of anticoagulant therapy for deep vein thrombosis and pulmonary embolism. Blood. 2014;123(12):1794-801.
13. Martinelli I, De Stefano V, Mannucci PM. Inherited risk factors for venous thromboembolism. Nat Rev Cardiol. 2014;11(3):140-56.
14. Boutitie F, Pinede L, Schulman S, Agnelli G, Raskob G, Julian J, et al. Influence of preceding length of anticoagulant treatment and initial presentation of venous thromboembolism on risk of recurrence after stopping treatment: analysis of individual participants' data from seven trials. BMJ. 2011;342:d3036.

15. Baglin T, Douketis J, Tosetto A, Marcucci M, Cushman M, Kyrle P, et al. Does the clinical presentation and extent of venous thrombosis predict likelihood and type of recurrence? A patient-level meta-analysis. J Thromb Haemost. 2010;8(11):2436-42.
16. Chee CE, Ashrani AA, Marks RS, Petterson TM, Bailey KR, Melton LJ, et al. Predictors of venous thromboembolism recurrence and bleeding among active cancer patients: a population-based cohort study. Blood. 2014;123(25):3972-8.
17. Hansson P, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: Incidence and risk factors. Arch Intern Med. 2000;160(6):769-74.
18. Schulman S, Granqvist S, Holmström M, Carlsson A, Lindmarker P, Nicol P, et al. The duration of oral anticoagulant therapy after a second episode of venous thromboembolism. N Engl J Med. 1997;336(6):393-8.
19. Eischer L, Eichinger S, Kyrle PA. The risk of recurrence in women with venous thromboembolism while using estrogens: a prospective cohort study. J Thromb Haemost. 2014;12(5):635-40.
20. Philbrick J, Shumate R, Siadaty M, Becker D. Air travel and venous thromboembolism: a systematic review. J Gen Intern Med. 2007;22(1):107-14.
21. Kearon C, Ginsberg JS, Hirsh J. The role of venous ultrasonography in the diagnosis of suspected deep venous thrombosis and pulmonary embolism. Ann Intern Med. 1998;129(12):1044-9.
22. Carrier M, Rodger MA, Wells PS, Righini M, Le Gal G. Residual vein obstruction to predict the risk of recurrent venous thromboembolism in patients with deep vein thrombosis: a systematic review and meta-analysis. J Thromb Haemost. 2011;9(6):1119-25.
23. Marcucci M, Smith CT, Douketis JD, Tosetto A, Baglin T, Cushman M, et al. Patient-level compared with study-level meta-analyses demonstrate consistency of D-dimer as predictor of venous thromboembolic recurrences. Journal of Clinical Epidemiology. 2013;66(4):415-25.
24. Palareti G, Cosmi B, Legnani C, Antonucci E, De Micheli V, Ghirarduzzi A, et al. D-dimer to guide the duration of anticoagulation in patients with venous thromboembolism: a management study. Blood. 2014;124(2):196-203.
25. Kearon C, Iorio A, Palareti G, on behalf of the subcommittee on control of anticoagulation of the SSC of the I. Risk of recurrent venous thromboembolism after stopping treatment in cohort studies: recommendation for acceptable rates and standardized reporting. J Thromb Haemost. 2010;8(10):2313-5.
26. Gómez-Outes A, Terleira-Fernández AI, Lecumberri R, Suárez-Gea ML, Vargas-Castrillón E. Direct oral anticoagulants in the treatment of acute venous thromboembolism: A systematic review and meta-analysis. Thromb Res. 2014;134(4):774-82.
27. Wu C, Alotaibi GS, Alsaleh K, Sean McMurtry M. Case fatality of bleeding and recurrent venous thromboembolism during, initial therapy with direct oral anticoagulants: A systematic review. Thromb Res. 2014;134(3):627-32.
28. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation. Chest. 2010;138(5):1093-100.
29. Schulman S, Rhedin A-S, Lindmarker P, Carlsson A, Lärfars G, Nicol P, et al. A comparison of six weeks with six months of oral anticoagulant therapy after a first episode of venous thromboembolism. N Engl J Med. 1995;332(25):1661-5.

30. Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, et al. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. N Engl J Med. 2001;345(3):165-9.
31. Pinede L, Ninet J, Duhaut P, Chabaud S, Demolombe-Rague S, Durieu I, et al. Comparison of 3 and 6 months of oral anticoagulant therapy after a first episode of proximal deep vein thrombosis or pulmonary embolism and comparison of 6 and 12 weeks of therapy after isolated calf deep vein thrombosis. Circulation. 2001;103(20):2453-60.
32. Agnelli G, Prandoni P, Becattini C, Silingardi M, Taliani MR, Miccio M, et al. Extended oral anticoagulant therapy after a first episode of pulmonary embolism. Ann Intern Med. 2003;139(1):19-25.
33. Campbell I, Bentley D, Presscott R, Routledge P, Shetty H, Williamson I. Anticoagulation for three versus six months in patients with deep vein thrombosis or pulmonary embolism, or both: randomised trial. BMJ. 2007;334(7595):674.
34. Kearon C, Gent M, Hirsh J, Weitz J, Kovacs MJ, Anderson DR, et al. A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med. 1999;340(12):901-7.
35. Farraj R. Anticoagulation period in idiopathic venous thromboembolism. How long is enough? Saudi Med J. 2004;25(7):848-51.
36. Palareti G, Cosmi B, Legnani C, Tosetto A, Brusi C, Iorio A, et al. D-dimer testing to determine the duration of anticoagulation therapy. N Engl J Med. 2006;355(17):1780-9.
37. Siragusa S, Malato A, Anastasio R, Cigna V, Milio G, Amato C, et al. Residual vein thrombosis to establish duration of anticoagulation after a first episode of deep vein thrombosis: the Duration of Anticoagulation based on Compression UltraSonography (DACUS) study. Blood. 2008;112(3):511-5.
38. Couturaud F, Sanchez O, Pernod G, et al. Six months vs extended oral anticoagulation after a first episode of pulmonary embolism: The padis-pe randomized clinical trial. JAMA. 2015;314(1):31-40.
39. Becattini C, Agnelli G, Schenone A, Eichinger S, Bucherini E, Silingardi M, et al. Aspirin for preventing the recurrence of venous thromboembolism. N Engl J Med. 2012;366(21):1959-67.
40. Brighton TA, Eikelboom JW, Mann K, Mister R, Gallus A, Ockelford P, et al. Low-dose aspirin for preventing recurrent venous thromboembolism. N Engl J Med. 2012;367(21):1979-87.
41. Rodger M, Kahn S, Wells P, Anderson D, Chagnon I, Le Gal G, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. CMAJ. 2008;179(5):417-26.
42. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation. 2010;121(14):1630-6.
43. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). J Thromb Haemost. 2012;10(6):1019-25.

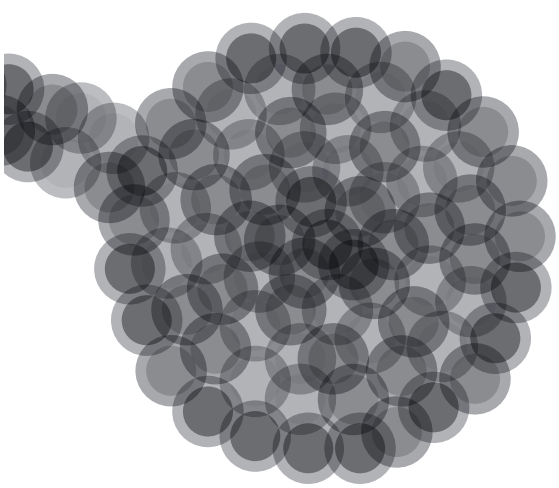


CHAPTER 7

VENOUS THROMBOEMBOLISM: TAILORING ANTICOAGULANT THERAPY DURATION (THE VISTA STUDY)

Janneke MT Hendriksen, Geert-Jan Geersing, Ruud Oudega,

Roger EG Schutgens, Karel GM Moons



Based on: Ned Tijdschr Hematol 2013;10:150-5.

SUMMARY

Introduction: Anticoagulant treatment with vitamin K antagonists (VKA) in patients with venous thromboembolism (VTE) aims at prevention of thrombus growth and reduction of the risk of recurrent VTE events. However, the major drawback of anticoagulation therapy is the concomitant risk of (major) bleeding. Consequently, the optimal duration of VKA therapy depends on an individual trade-off between the risks of VTE recurrence and bleeding. Especially in patients with unprovoked VTE (that is: VTE in the absence of a major transient risk factor and a consequently higher recurrence risk) it is discussed if six months of anticoagulant treatment is sufficient, or should be extended for an indefinite period. Recent studies suggest that the optimal duration for each individual patient with unprovoked VTE can be estimated by the use of a prediction model for the risk of recurrent VTE; prolonged or indefinite treatment would then be advocated if this predicted recurrence risk is above a threshold of 5% per year. The efficacy and safety of such a tailored treatment strategy based on a prediction model have never been prospectively investigated before.

Objectives: In patients with unprovoked VTE, we compare the effect of prediction model guided VKA treatment duration compared to standard treatment duration ('care-as-usual'), in terms of incidence of VTE recurrence during 24 months of follow-up and consequent cost-effectiveness.

Patients and methods: Consecutive patients with an unprovoked VTE receiving anticoagulant treatment by one of the participating Thrombosis Services, are randomized to 'care as usual' or the intervention: a tailored treatment duration guided by the Vienna prediction model (combining gender, VTE localization and D-dimer testing) after the initial treatment period of 6 months. If the prediction model yields a low risk of VTE recurrence (<5 % recurrence in first year), six months of treatment is recommended. In case of a high recurrence risk (≥ 5% recurrence in first year) treatment is recommended to be continued for another 24 months. The primary endpoint is a symptomatic recurrent VTE event (proximal DVT and fatal or non-fatal pulmonary embolism) during 24 months of follow-up. The secondary outcomes include major bleeding events, quality of life and cost-effectiveness.

Patients in both groups undergo D-dimer testing twice: six months after the initial VTE, and 28 days after VKA withdrawal. Venous plasma will be stored for further evaluation after completion of the study. To be able to demonstrate a reduction of the number of recurrent VTE events in the intervention group from 7% to 3.5% per year of follow-up, 692 patients are needed per treatment arm (two-sided alpha 0.05, power 80%). Accounting for loss-to-follow-up, we aim to include 750 patients per arm of the trial.

BACKGROUND

Patients with venous thromboembolic disease (VTE, including deep venous thrombosis (DVT) and pulmonary embolism (PE)) are treated with oral anticoagulants (usually vitamin K antagonists (VKA)) to prevent thrombus growth and recurrent events. The risk of VTE recurrence largely determines the duration of anticoagulant therapy. In the presence of a major transient risk factor like recent surgery or immobilisation, treatment duration is confined to a period of three months due to the low risk of VTE recurrence (3% in the first year). If such transient risk factors are absent (unprovoked VTE), the risk of VTE recurrence is much higher (30-40% within 5 years after VKA withdrawal). (1) Consequently, current guidelines advice anticoagulant treatment for (at least) 6 months. (2, 3)

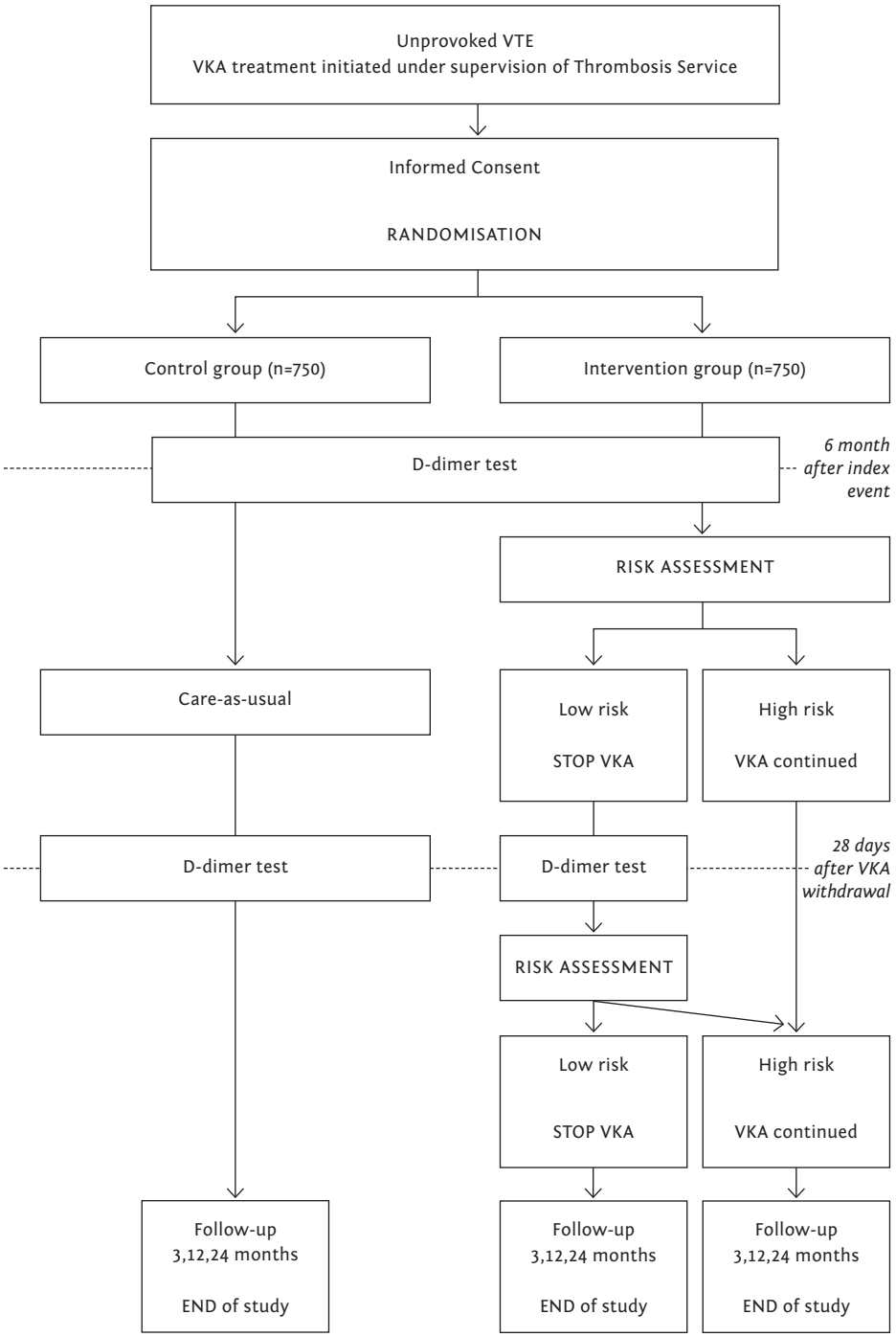
However, prolonged anticoagulant treatment beyond six months after a first event is thought to be beneficial in a subset of patients at high risk of recurrence. In those cases, prevention of a recurrent event and the subsequent lower risk of chronic long-term complications of VTE like post-thrombotic syndrome and chronic thromboembolic pulmonary hypertension (4, 5) outweigh the risk of bleeding while on anticoagulant treatment. Current guidelines leave this treatment decision to the physician's own discretion, which causes lack of clarity and a wide variety in clinical practice. (2, 3, 6) Although several studies have demonstrated the effectiveness of prolonged anticoagulation therapy after a first unprovoked VTE, (7-10) its implementation is hampered by the fact that the individual balance between bleeding and thrombotic risk has not been taken into account.

Therefore, anticoagulation therapy could benefit from tailored therapy based on individual risk assessment. Previous studies demonstrated that clinical variables like elevated D-dimer levels after VKA withdrawal, residual vein thrombosis after a DVT and male gender are associated with an increased risk of recurrence. (11-14) Recently, several prediction models for recurrent VTE- such as the Vienna prediction model- have been developed combining several of these potential predictor variables into a score that can be used to facilitate risk stratification. (15-17) If the model identifies a low risk of recurrence (‹that is: a one-year recurrence risk of ‹5%), (18) six months of anticoagulant therapy is likely to be enough, whereas longer treatment duration may be warranted in those at high risk of recurrence (that is: ≥ 5% recurrence risk in first year). However, a prospective study to assess the efficacy and safety of such a tailored treatment based on a risk prediction, compared to current usual care, has not been performed yet.

STUDY AIM

In patients with unprovoked VTE, we compare the effect of prediction model guided VKA treatment duration compared to standard treatment duration ('care-as-usual'), in terms of incidence of VTE recurrence during 24 months of follow-up and consequent cost-effectiveness.

FIGURE 1 FLOW CHART VISTA TRIAL



METHODS

STUDY DESIGN

The VISTA study (Venous thrombo-embolism: tailoring anticoagulant therapy duration) is a pragmatic randomized intervention study. Participant recruitment takes place in close collaboration with multiple Thrombosis Services in The Netherlands. After written informed consent, participants are randomized to a control or intervention group. Randomization is blocked by Thrombosis Service and the type of the index event (DVT or PE) and is performed within six months of initial anticoagulant treatment, preferably at about five months after initiation of anticoagulation treatment. See *Figure 1* for the study flow chart.

POPULATION

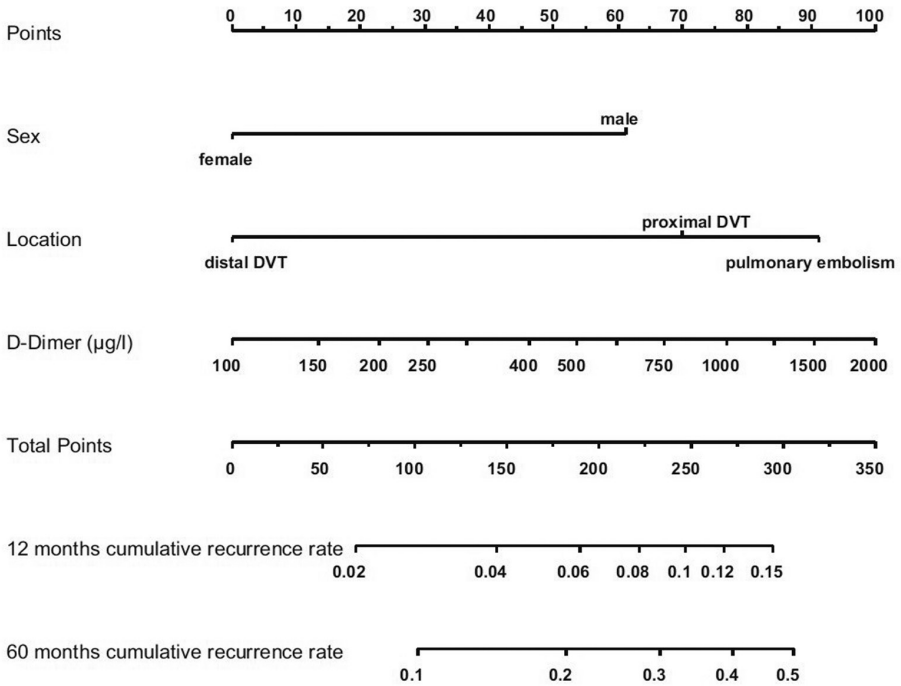
All adult subsequent patients, newly registered by one of the participating Thrombosis Services with the indication for an anticoagulant treatment period of at least three to six months due to an unprovoked VTE, are invited to participate. Excluded are all patients with VTE episodes preceded by a recent (< 3 months) surgery, immobilization by cast, (active treatment for) malignancy, pregnancy/puerperium, or in case of a known antiphospholipid antibody syndrome. Also, a previous venous thromboembolic event should not have occurred within 10 years prior to the current index event. The treating physician, that is internal medicine specialist, pulmonologist and/or general practitioner, will be notified on the participation of their patient in this study.

INTERVENTION

In the intervention group patients undergo a formal risk assessment by applying the Vienna prediction model developed by Eichinger et al. (15) at the end of their initial six-month treatment period. This prediction model has been developed in a cohort of 979 patients with a first unprovoked VTE who have been followed up for a median of 43.3 months after VKA withdrawal for recurrence of VTE. A total of 176 patients (18.9%) had a recurrent event. From a preselected set of variables that were thought to be associated with VTE recurrence, the variables male gender (versus female gender hazard ratio (HR) 1.90 (95% confidence interval (CI) 1.31 to 2.75), localization of the thrombus (PE versus distal DVT HR 2.60 (95% CI 1.49 to 4.53) and D-dimer level (each doubling HR 1.27 (95% CI 1,08 to 1,51) appeared to be predictive for recurrence of VTE in a Cox proportional hazards model. The model can be applied using a nomogram or overview (see *Figure 2* and *Table 1*) and is online available as a web-based tool (http://cemsiis.meduniwien.ac.at/en/kb/science-research/software/clinical-software/recurrent-vte/).

To enable a risk assessment using this model, all patients undergo a blood withdrawal for D-dimer testing. First, the D-dimer level will be determined while still using anticoagulants at the end of the initial six-month treatment period. If the prediction model estimates a one-year VTE recurrence risk <5% at six months

FIGURE 2 NOMOGRAM OF THE VIENNA PREDICTION MODEL (15)



Based on: Eichinger et al. Circulation. 2010;121(14):1630-6.

of anticoagulant treatment, the patient is advised to discontinue anticoagulant treatment after the initial six months of treatment. A second blood draw will be performed 28 days after VKA withdrawal, to intercept a potential increase of D-dimer level due to the withdrawal of VKA. (19) In case the model predicts an elevated recurrence risk (that is: ≥ 5 % VTE recurrence risk during the first year after VKA withdrawal) in one of the two risk assessments, the VKA treatment is advised to be continued or resumed for a prolonged period of another 24 months. This treatment recommendation is performed in close collaboration with patient preference and can always be overruled by the treating physician if he deems the risk of bleeding too high, considers another contra-indication for continued treatment to be present or thinks the risk of recurrence is too high to stop anticoagulation treatment (e.g. a young female with hemodynamically unstable massive pulmonary embolism in need for thrombolysis).

CARE-AS-USUAL

The strategy as described above will be compared to usual care in the control group. Care-as-usual implies a treatment duration at the physician's own discretion, based on current guidelines. This usually implies 6 months of treatment.

TABLE 1 PREDICTED 1-YEAR RISKS OF RECURRENT VTE, BASED ON THE VIENNA PREDICTION MODEL (15)

	Female			Male		
D-dimer (µg/l)	**Distal DVT**	**Proximal DVT**	**PE**	**Distal DVT**	**Proximal DVT**	**PE**
100	1.1 (0.6-1.8)	2.0 (1.1-3.7)	2.5 (1.4-4.4)	1.9 (1.0-3.5)	3.6 (2.1-6.0)	4.3 (2.7-6.9)
200	1.3 (0.7-2.4)	2.5 (1.4-4.5)	3.1 (1.8-5.3)	2.3 (1.3-4.2)	4.4 (2.7-7.0)	5.3 (3.5-8.1)
300	1.5 (0.8-2.7)	2.9 (1.6-5.0)	3.5 (2.0-5.8)	2.6 (1.5-4.7)	5.0 (3.2-7.7)	6.0 (4.0-9.0)
400	1.6 (0.9-3.0)	3.1 (1.8-5.4)	3.8 (2.3-6.2)	2.9 (1.6-5.0)	5.4 (3.5-8.2)	6.5 (4.4-9.6)
500	1.8 (1.0-3.2)	3.3 (2.0-5.7)	4.0 (2.5-6.6)	3.1 (1.8-5.3)	5.8 (3.8-8.7)	7.0 (4.8-10.2)
600	1.9 (1.0-3.4)	3.5 (2.1-5.9)	4.3 (2.6-6.8)	3.3 (1.9-5.5)	6.1 (4.1-9.1)	7.4 (5.1-10.7)
700	2.0 (1.1-3.6)	3.7 (2.2-6.1)	4.5 (2.8-7.1)	3.4 (2.0-5.8)	6.4 (4.3-9.4)	7.7 (5.3-11.1)
800	2.0 (1.1-3.7)	3.8 (2.3-6.3)	4.6 (2.9-7.3)	3.5 (2.1-5.9)	6.7 (4.5-9.8)	8.0 (5.6-11.5)
900	2.1 (1.1-3.8)	4.0 (2.4-6.5)	4.8 (3.1-7.5)	3.7 (2.2-6.1)	6.9 (4.7-10.1)	8.3 (5.8-11.9)
1000	2.2 (1.2-4.0)	4.1 (2.5-6.7)	5.0 (3.2-7.7)	3.8 (2.3-6.3)	7.1 (4.9-10.3)	8.6 (6.0-12.2)
1100	2.2 (1.2-4.1)	4.2 (2.6-6.8)	5.1 (3.3-7.9)	3.9 (2.4-6.4)	7.3 (5.0-10.6)	8.8 (6.2-12.6)
1200	2.3 (1.3-4.2)	4.3 (2.7-6.9)	5.2 (3.4-8.0)	4.0 (2.5-6.5)	7.5 (5.2-10.8)	9.1 (6.3-12.9)
1300	2.4 (1.3-4.3)	4.4 (2.8-7.1)	5.4 (3.5-8.2)	4.1 (2.5-6.6)	7.7 (5.3-11.1)	9.3 (3.5-13.2)
1400	2.4 (1.3-4.3)	4.5 (2.9-7.2)	5.5 (3.6-8.3)	4.2 (2.6-6.8)	7.9 (5.5-11.3)	9.5 (6.6-13.5)
1500	2.5 (1.4-4.4)	4.6 (2.9-7.3)	5.6 (3.7-8.5)	4.3 (2.7-6.9)	8.0 (5.6-11.5)	9.7 (6.7-13.7)
1600	2.5 (1.4-4.5)	4.7 (3.0-7.4)	5.7 (3.8-8.6)	4.4 (2.7-7.0)	8.2 (5.7-11.7)	9.8 (6.9-14.0)
1700	2.6 (1.4-4.6)	4.8 (3.1-7.5)	5.8 (3.9-8.7)	4.5 (2.8-7.1)	8.3 (5.8-11.9)	10.0 (7.0-14.3)
1800	2.6 (1.5-4.6)	4.9 (3.1-7.6)	5.9 (3.9-8.9)	4.5 (2.9-7.2)	8.5 (5.9-12.1)	10.2 (7.1-14.5)
1900	2.6 (1.5-4.7)	5.0 (3.2-7.7)	6.0 (4.0-9.0)	4.6 (2.9-7.3)	8.6 (6.0-12.3)	10.4 (7.2-14.7)
2000	2.7 (1.5-4.8)	5.1 (3.3-7.8)	6.1 (4.1-9.1)	4.7 (3.0-7.3)	8.7 (6.1-12.4)	10.5 (7.3-15.0)

DVT= deep venous thrombosis; PE= pulmonary embolism

OUTCOME

The primary outcome is a symptomatic recurrent VTE (DVT and (non-) fatal pulmonary embolism) during the 24-month follow-up. Recurrent thrombosis is diagnosed according to current diagnostic standards (new thrombus in venous segment or extension of a persisting thrombus) and treated according to the standard therapeutic protocols. The secondary outcomes include major bleedings, as defined by the ISTH criteria, quality of life and cost-effectiveness over the same follow-up period. (20) Thrombosis Services and treating physicians are asked to inform the researchers in case of a recurrent VTE or bleeding event. Furthermore, all patients are contacted by telephone three times during follow-up by one of the research nurses.

DATA ANALYSIS

SAMPLE SIZE CALCULATION

The incidence of recurrent VTE varies in observational studies from 6-9% within 12 months after VKA discontinuation in unprovoked VTE. (11, 21) In a randomized controlled trial on a tailored VKA duration based on D-dimer levels in patients with unprovoked VTE, the recurrent rate at 18 months of follow-up was 15.0% in patients with positive D-dimer levels randomized to VKA discontinuation and 2.9% in patients randomized to prolongation of VKA. (22) Recalculating these numbers and if care-as-usual was to be applied to all patients of this trial, an overall recurrence rate of 9% would be expected. This rate would be 3.4% if all patients with a positive D-dimer test had been treated. Observational studies show that the VTE recurrence rate in D-dimer negative patients is 3.5%. (11) Based on these numbers, we conservatively assume a VTE recurrence rate of 7% in the control group and 3.5% in the intervention group per year. To determine whether the intervention will lower the recurrence rate from 7% to 3.5%, with a power of 80% and a two-sided alpha of 5%, at least 692 patients per arm will be needed. Taking into account a loss of follow-up of 10%, we aim to include 750 patients per study arm.

STATISTICAL ANALYSES

The aim of the main analysis is to compare the incidence of the primary outcome (VTE recurrence within 24 months after the intervention) in both groups. The results will be presented as relative risk with a surrounding 95% confidence interval. The risk difference (with 95% confidence interval) and the number needed to treat will be estimated. These analyses will also be performed for the secondary outcome measure major bleeding.

Then, Cox survival analysis will be used to compare the time to recurrent VTE events of both groups, adjusted for center and type of VTE (DVT and PE). All analysis will be done according to the intention to treat principle but will be assessed per protocol as well. Finally, for the cost-effectiveness analyses a Markov model will be built with the data obtained prospectively on costs and effects. The influence of the parameter uncertainty will be further tested using probabilistic sensitivity analysis using Monte Carlo simulations.

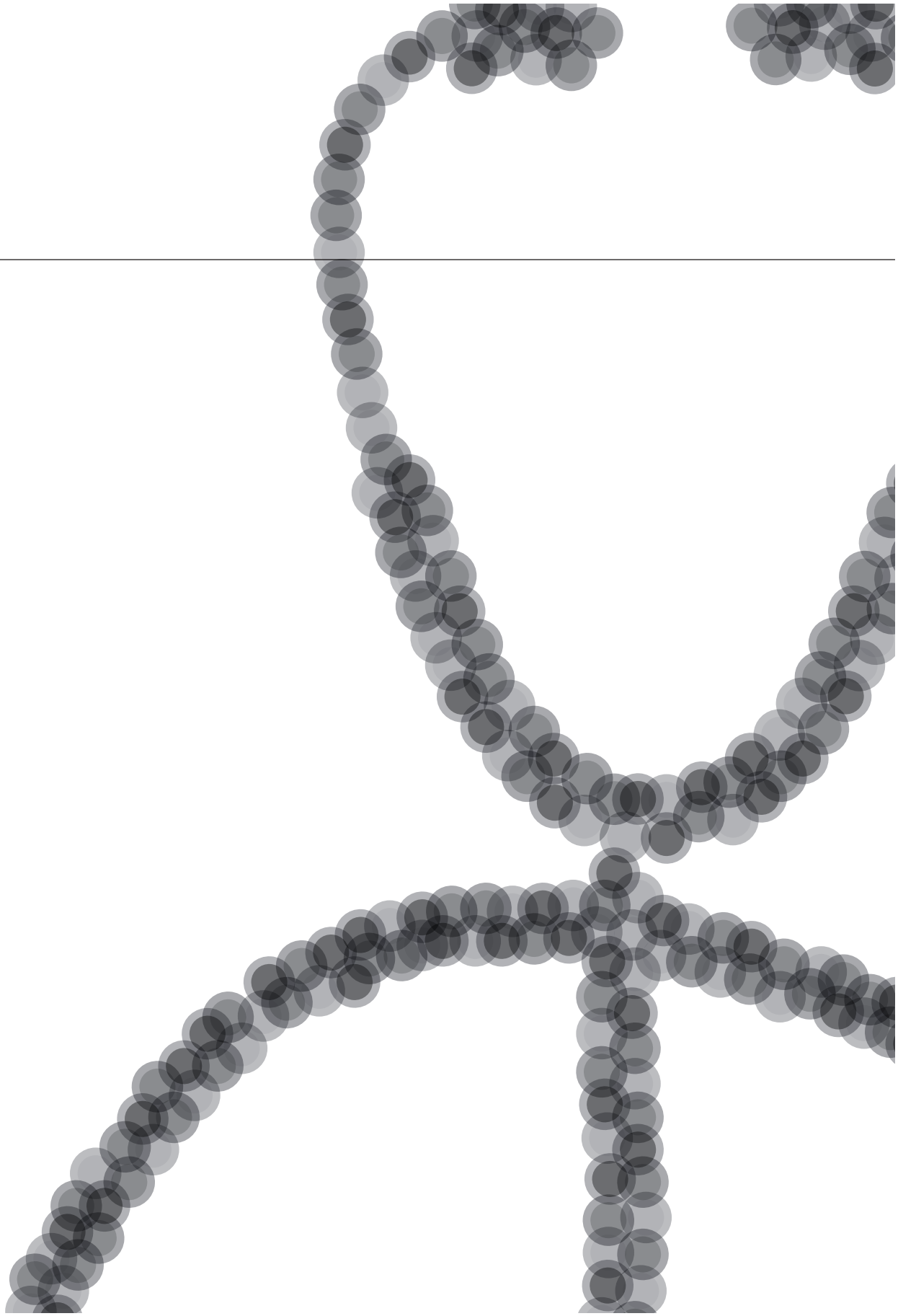
CURRENT STATUS

In September 2011 the first patients have been included, the first intervention has taken place in January 2012. After starting at five Thrombosis Services, four other Services joined during the course of 2013 and 2014. Inclusion will close in the course of 2015.

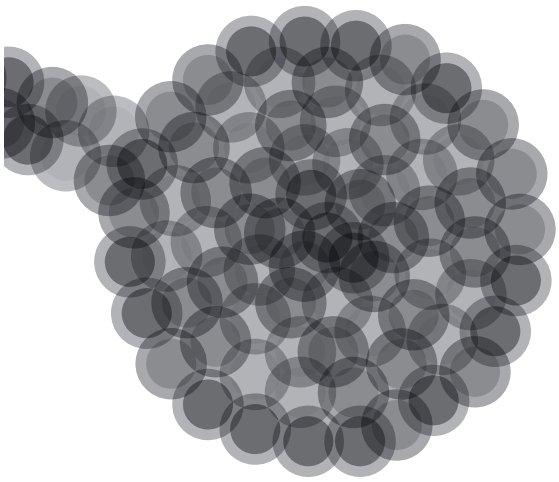
REFERENCES

1. Prandoni P, Lensing AW, Cogo A, Cuppini S, Villalta S, Carta M, et al. The long-term clinical course of acute deep venous thrombosis. Ann Intern Med. 1996;125(1):1-7.
2. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-based clinical practice guidelines. Chest. 2012;141(2_suppl):e419S-e94S.
3. Baaten G, van Bakel H, Geersing G, de Jong J, Lucassen W, van Noortwijk H, et al. Diepe veneuze trombose en longembolie (eerste herziening). Huisarts Wet. 2015;58(1):26-35.
4. de Wolf M, Wittens C, Kahn SR. Incidence and risk factors of the post-thrombotic syndrome. Phlebology. 2012;27(Suppl 1):85-94.
5. Bonderman D, Wilkens H, Wakounig S, Schäfers HJ, Jansa P, Lindner J, et al. Risk factors for chronic thromboembolic pulmonary hypertension. Eur Respir J. 2009;33(2):325-31.
6. Kwaliteitsinstituut voor de Gezondheidszorg CBO. Richtlijn: Diagnostiek, preventie en behandeling van veneuze trombo-embolie en secundaire preventie arteriele trombose. Utrecht 2008.
7. Ridker PM, Goldhaber SZ, Danielson E, Rosenberg Y, Eby CS, Deitcher SR, et al. Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. N Engl J Med. 2003;348(15):1425-34.
8. Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, et al. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. N Engl J Med. 2001;345(3):165-9.
9. Kearon C, Gent M, Hirsh J, Weitz J, Kovacs MJ, Anderson DR, et al. A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med. 1999;340(12):901-7.
10. Agnelli G, Prandoni P, Becattini C, Silingardi M, Taliani MR, Miccio M, et al. Extended oral anticoagulant therapy after a first episode of pulmonary embolism. Ann Intern Med. 2003;139(1):19-25.
11. Verhovsek M, Douketis JD, Yi Q, Shrivastava S, Tait RC, Baglin T, et al. Systematic review: D-dimer to predict recurrent disease after stopping anticoagulant therapy for unprovoked venous thromboembolism. Ann Intern Med. 2008;149(7):481-90.
12. Cosmi B, Legnani C, Iorio A, Pengo V, Ghirarduzzi A, Testa S, et al. Residual venous obstruction, alone and in combination with D-dimer, as a risk factor for recurrence after anticoagulation withdrawal following a first idiopathic deep vein thrombosis in the Prolong study. Eur J Vasc Surg. 2010;39(3):356.
13. Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, et al. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. BMJ. 2011;342.
14. McRae S, Tran H, Schulman S, Ginsberg J, Kearon C. Effect of patient's sex on risk of recurrent venous thromboembolism: a meta-analysis. Lancet. 2006;368(9533):371-8.
15. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation. 2010;121(14):1630-6.

16. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). J Thromb Haemost. 2012;10(6):1019-25.
17. Rodger M, Kahn S, Wells P, Anderson D, Chagnon I, Le Gal G, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. CMAJ. 2008;179(5):417-26.
18. Kearon C, Iorio A, Palareti G, on behalf of the subcommittee on control of anticoagulation of the SSC of the I. Risk of recurrent venous thromboembolism after stopping treatment in cohort studies: recommendation for acceptable rates and standardized reporting. J Thromb Haemost. 2010;8(10):2313-5.
19. Ombandza-Moussa E, Samama M, Horellou M, Elalamy I, Conard J. Potential use of D-dimer measurement in patients treated with oral anticoagulant for a venous thromboembolic episode. Int Angiol. 2003;22(4):364-9.
20. Schulman S, Kearon C, the SOCOAOTS, Standardization Committee Of The International Society On T, Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost. 2005;3(4):692-4.
21. Baglin T, Palmer CR, Luddington R, Baglin C. Unprovoked recurrent venous thrombosis: prediction by D-dimer and clinical risk factors. J Thromb Haemost. 2008;6(4):577-82.
22. Palareti G, Cosmi B, Legnani C, Tosetto A, Brusi C, Iorio A, et al. D-dimer testing to determine the duration of anticoagulation therapy. N Engl J Med. 2006;355(17):1780-9.



GENERAL DISCUSSION



MAIN CONCLUSIONS

In this thesis, we evaluated diagnostic and prognostic approaches in patients with suspected venous thromboembolism (VTE) in primary care, aiming to improve the timely identification of VTE cases while minimizing over-referral, and to improve individually tailored treatment.

- Delay of more than 7 days between first presentation at the general practitioner (GP) and the diagnosis of pulmonary embolism (PE) is observed in 26% of the PE cases. Older age and absence of chest complaints like pain on inspiration are determinants associated with diagnostic delay.
- To improve the diagnostic (and prognostic) approaches in thromboembolic disease, researchers should better adhere to the suggested phases in diagnostic and prognostic prediction modeling.
- Of all diagnostic prediction models developed for suspected pulmonary embolism, the Wells PE rules (original, modified and simplified) perform best in primary care.
- Although both are safe approaches, the Wells PE rule is more efficient than the GP's own gestalt estimate, to rule out pulmonary embolism in suspected patients in primary care.
- In the diagnostic process of DVT in primary care, the use of point-of-care D-dimer tests results in comparable long-term health outcomes compared to laboratory based testing procedures, but can be performed more easily and at lower costs.
- Prevention of recurrent VTE in patients with an unprovoked event is expected to improve with the introduction of prognostic prediction models for recurrent VTE. Prolonged duration of anticoagulation therapy can be confined to those patients with a predicted increased risk of VTE recurrence.
- The actual use of a prognostic prediction model to tailor treatment duration in patients with unprovoked VTE is currently under evaluation in a large scale pragmatic randomized trial.

In this final chapter, the clinical implications and future perspectives of diagnostic and prognostic risk stratification of venous thromboembolism are discussed, with a focus on primary care medicine.

DIAGNOSIS OF VTE IN PRIMARY CARE

1. CLINICAL IMPLICATIONS

In January 2015, the revised version of the Dutch primary care guideline on VTE diagnosis and management was published. (1) The Wells PE rule has been implemented in this updated guideline, based on the rule's good performance in the primary care validation study "AMUSE-2". (1, 2) GPs are encouraged to use the Wells PE rule in patients suspected of PE, with the aim to decrease the number of

referrals to secondary care while not compromising on safety.

Whether or not these aims will be met in daily care depends on several factors. One obvious factor is the accuracy of the model itself in a 'real world setting'. Two other important factors are 1) when (that is: in which patients) to use the diagnostic prediction model, and 2) the potential drawbacks to use a diagnostic prediction model in clinical practice.

When to use a diagnostic prediction model?

Adequate identification of venous thromboembolism obviously starts with taking the diagnosis into consideration. The whole diagnostic process and associated steps as described in Chapter 2-5 are based on the fact that pulmonary embolism is at least suspected by the GP. However, in Chapter 1 we demonstrated that PE is not suspected at initial presentation in about one quarter of all PE cases. Consequently, this group of patients is (most likely) not included in all studies aimed at validating the accuracy of diagnostic prediction models in primary care. As such, increased awareness by GPs on the fact that pulmonary embolism can present in a very nonspecific manner can have effects on the estimated diagnostic accuracy of these prediction models. Ideally, increased awareness leads to broader suspicion of patients presenting at their general practitioner, an increase of correct and direct PE diagnoses, but without loss of efficiency when using the prediction model. However, the risk of an overshoot of referrals if all patients with nonspecific complains are evaluated for pulmonary embolism in primary care is reasonable. In case of a low probability based on the Wells rule, a D-dimer test has to be performed. And especially in the elderly or those with comorbidity who often present with non-specific symptoms, the D-dimer result is often false- positive. (3-5) In this scenario, the lowering of the threshold to consider the diagnosis would further hamper the performance of the prediction model. There is a chance that broadening the suspicion of PE by GPs leads to decreased efficiency of the use of prediction models, due to the number of unnecessary referrals of eventually non-PE cases who presented with a specific symptoms and signs. If this overshoot of referrals should be accepted in an attempt to prevent missing diagnosis, is largely dependent on the clinical consequences of a delayed or missed PE diagnosis.

Based on the literature on pulmonary embolism treatment in the beginning of the 1960s up to the 1990s, the mortality figures of untreated pulmonary embolism have been estimated to be between 10-30%. (6, 7) In contrast, the in-hospital mortality rate of VTE patients with immediately initiated anticoagulant treatment is reported to be 1.5% (95% CI 0.9% to 2.2%). (8) From this point of view, prompt diagnosis of VTE, and subsequent treatment initiation, is of uttermost importance and recommended by current international guidelines. (9, 10)

However, little is known about the impact of diagnostic delay on mortality, morbidity, quality of life and costs involved, in the light of current improved diagnostic and therapeutic options. Delay was not associated with increased

mortality in the first three months post-event in three studies on diagnostic delay of PE in emergency departments, however only patients who survived this period of delay were included. (11-13)

It is difficult to design a study that accurately evaluates delay, and the consequences of missing or delayed diagnoses of PE. Only in retrospect, delay can be labeled as such. In addition, these studies can only be performed in those patients surviving the period of delay. The subset of patients that dies due to a missed diagnosis will only be identified as such if post-mortem information is available.

Prospective follow-up of all patients presenting at their GP with the broadest range of PE related signs and symptoms, will imply the inclusion of many patients. Furthermore, there needs to be a (gold) reference standard in all to ultimately confirm or reject the PE diagnosis. With the practical difficulties arising with a prospective research design, an alternative (retrospective) approach requires adequate classification (ICPC coding) and extensive documentation, in order to be able to perform a study on determinants on diagnostic delay, as soon as possible after the diagnosis PE. In these (later or delayed) identified PE cases, further prospective evaluation is needed to be able to assess long-term outcomes and to determine whether these were indeed mild PE cases after all.

The results of such studies might gain more insight in the clinical consequences of delayed diagnoses of PE, and the need to act upon this in the diagnostic process of the GP. One can think of increasing the threshold to refer a patient if the effects of delay are not as large as currently considered, or accept over-referral if consequences of a missed or delayed diagnosis are considered to be substantial.

Potential drawbacks to use a diagnostic prediction model

Despite the fact that the current guidelines encourage general practitioners to use the Wells diagnostic prediction model in suspected patients, it is thought that some GPs deviate from the guidelines nevertheless and rely on their own gestalt estimate instead. However, no exact numbers on the use of gestalt versus diagnostic prediction models for suspected PE are available. In a French study, general practitioners were interviewed on their diagnostic strategies in pulmonary embolism cases. None of the general practitioners used a prediction model to guide their decisions. (14) Also, in our study on diagnostic delay (Chapter 1) the chart reviews fueled the question if the Wells prediction model and D-dimer testing were used on a structural basis. For instance, the result of the Wells rule was rarely documented in the records of the patients who consulted their general practitioner with complaints suggestive for pulmonary embolism. Furthermore, D-dimer levels were documented in only 14% of pulmonary embolism cases.

This limited documentation of Wells rule and D-dimer results can be related to the fact that the use of a prediction model and D-dimer testing in suspected pulmonary embolism were introduced to the guidelines only recently. Another explanation might be limited, or incorrect, use of the prediction model in daily

clinical practice. When discussing the use of prediction models with general practitioners, obstacles to use the diagnostic prediction models are often mentioned. These obstacles include the conceived "hassle" to use a prediction model during a busy clinic, no readily available D-dimer testing, prior negative experience with using the prediction model and/ or D-dimer test, and (a fear to be confronted with) incongruence in the own gestalt estimation and the results of the prediction model.

"**A hassle**": The Wells PE rule consists of 7 items that have to be scored. (15) Different weights (0-1-1.5-3) are assigned to different items of the rule. As not only the weights per item, but also the items in itself are easily forgotten, it takes time to determine the exact score for a patient. Furthermore, computing errors are easily made, leading to an incorrect probability estimate and consequent wrong diagnostic decisions. However, these issues can be dealt with rather easily, for example by using an app, of using the simplified version of the Wells rule (see Chapter 3). (16)

Incorrect use: In specific groups of patients suspected of PE, the Wells rule should not be used, as the rule was not developed nor validated in these specific patients. This is for example the case for pregnant women, those with a suspected recurrent VTE and those already on anticoagulants. GPs need to be made more aware of these limitations of using the Wells PE rule, in order to prevent diagnostic errors.

Availability & test characteristics D-dimer testing: If the diagnostic prediction model predicts a low PE probability, a D-dimer test should be performed next. This laboratory test can be performed in a central lab facility. However this can lead to complex logistic procedures, especially if the lab facility is not nearby, or if the consultation takes place during out-of-office hours. If any of these obstacles is foreseen, some GPs choose to refer a patient without the result of a D-dimer anyhow. Others postpone testing until the next day, thereby accepting the risk of disease progression in the meanwhile.

An alternative to sending a patients to a central laboratory facility might be the use of point-of-care tests (POCTs) in primary care. When solely looking at the convenience of using a lab test at the office, the advantages of using POCTs are clear. However, the POCTs currently available have their drawbacks, in terms of a lower sensitivity, and for the dichotomous POCTs also a lack of quantitative results and even non-conclusive test results. (17) Each of these factors can cause general practitioners not to favor using POCTs and refer patients for laboratory based testing instead. In order to improve the conditions to use POCTs in primary care, one can think of frequent training of staff to prevent pre-analytical errors, or switching to one of the more sensitive quantitative POCTs, as described in Chapter 5.

Prior negative experiences with using a prediction model and/ or D-dimer test: Although the combination of the Wells PE rule and D-dimer testing has been shown to safely rule-out PE in about 4 out of 10 suspected patients, there is always the – albeit small- chance of a false-negative diagnostic classification by the model.

Such an experience can seriously hamper ones confidence in using a structured prediction model in daily practice.

Incongruence in the own (gestalt) estimation and the results of the prediction model: Especially in a primary care setting with only limited diagnostic tools available, it is extremely difficult for general practitioners to rule-out pulmonary embolism for sure. The GP's "gut feeling" or so called gestalt estimate is an extremely important diagnostic tool, but not as efficient as a structured probability estimate based on a prediction model: in Chapter 4, we demonstrated that the use of the Wells rule was just as safe as a GPs own gestalt estimate and much more efficient, at least in a study setting. As such, a diagnostic prediction model can help a GP to base the decision not to refer a patient. However, GPs can (and should) not ignore their gut feeling and are likely to overrule the outcome of the prediction model if there is a continuous sense of discomfort: the patient will be referred despite the low predicted PE probability.

All these factors might influence the actual performance of a diagnostic prediction model in primary care. Results of our studies on the use of alternative prediction models, gestalt and POC D-dimer testing can endorse the use of prediction models in primary care. However, the next important step is to evaluate the performance in a 'real life setting', in order to identify targets for further improvement of diagnostic strategies.

II. FUTURE PERSPECTIVES

As the Wells PE rule has been integrated in the Dutch GP guidelines (1), the focus will shift towards the impact of these rules on daily practice. As has been mentioned in the second chapter of this thesis, it is very important to continue evaluation of newly developed or externally validated prediction models in the setting in which the model is introduced. The potential problems that have been described in the previous section on the clinical implications can seriously hamper the ultimate value of this prediction model in daily clinical care.

Currently, two studies are running on the impact of actually using the Oudega DVT rule and Wells PE rule in primary care. The so-called Advice DVT and PE studies are intended to quantify the effects of the implementation of the VTE diagnostic prediction models in the Dutch primary care guidelines. The results are expected to be available in the course of 2016. Insight into barriers to use these models and mistakes commonly made in the use of these models, can be used to improve its value in the diagnostic process in primary care.

However, even with further fine-tuning of the use of diagnostic prediction models in primary care, based on the currently running impact studies, there is room for further improvement for risk stratification in primary care. The GPs diagnostic starting point is often different from the starting point of using a prediction model: a general practitioner is confronted with a patient presenting with a

certain subset of complains rather than presenting with "a suspected diagnosis". For example, a GP will only use a PE prediction model if he/she acknowledges the symptom unexplained coughing as a possible symptom of PE. And in case of chest pain, both a model for the diagnosis of an acute coronary event and a PE model could be relevant to use. It is not about the diagnosis pulmonary embolism yes or no, but about the probability of a severe disease being present that needs further action, usually referral to secondary care. Thus, a prediction model that takes into account multiple diagnoses as outcome measure would be a more practical (or logical) approach in a primary care setting. Especially in the elderly, in whom multi-morbidity is common, this multiple-disease approach might be very helpful. One might think of a ranking of disease probabilities, or a (less disease specific) mortality or morbidity risk based on the integration of signs and symptoms, with referral to secondary care if a certain threshold is passed. As such, it is not the suspicion of a distinct disease, but rather the specific combination of signs and complaints, that guides further diagnostic or therapeutic steps.

PROGNOSIS AND TREATMENT OF VTE IN PRIMARY CARE

FUTURE PERSPECTIVES

While the results of validation studies of prognostic prediction models for recurrent VTE, like VISTA, are eagerly awaited for, one can already think about the future perspectives of prognostic risk stratification in venous thromboembolic disease in primary care. Four aspects deserve further considerations:

1. Alternative biomarkers for VTE recurrence risk prediction

The VTE recurrence prognostic prediction models currently developed include the concentration of the fibrin degradation product D-dimer. The result of such tests, however, is heavily influenced by use of the anticoagulant therapy with vitamin K antagonists (VKA). (18, 19) Anticoagulant therapy needs to be stopped for a period of several weeks before a D-dimer test can validly be used in the prognostic model. (20-22) This can seriously hamper the use of these models in clinical practice. During this period of withdrawal and restart of medication, a patient is prone to a recurrent VTE event. (20, 23) Furthermore, in case of a high predicted risk, the anticoagulant therapy needs to be restarted. While the VKA dosing is adjusted until therapeutic dosing is re-achieved, a patients is at increased risk of bleeding. (24, 25)

Therefore, it is important to search for other biomarkers that are independent of the intensity of anticoagulant therapy. Platelet activation markers are thought to provide information on the thrombotic potential of a patient, but are expected not to be directly influenced by VKA. (26, 27) Further studies on this potential role in VTE recurrence prediction are planned.

2. Introduction DOACs

Over the last 5-8 years, there has been a tremendous evolution in the treatment of both arterial and venous thromboembolism. Whereas the use of VKA, combined with initial treatment with low molecular weight heparins until the therapeutic dose of VKA has been reached, has been the standard treatment for decades, the direct oral anticoagulants (DOACs) have been introduced as an alternative to VKA. (28-32) The facts that no INR measurement is needed anymore, the fixed dose regimen and the observed non-interiority over VKA in VTE treatment, with a tendency towards a reduced intra-cranial bleeding risk, contribute to the excitement over this new class of anticoagulants.

However, it should be stressed that the experience with these new agents is still limited to predominantly strictly controlled trial settings. Those patients most likely to be the largest part of the new consumers, namely elderly (indication AF, VTE) and those with multiple co-existing conditions, were excluded in many of the trials. Long-term experience is lacking until now.

Furthermore, a strong risk factor for bleeding is fluctuation of INR levels, expressed as a low time within therapeutic range (TTR). (33) The Dutch structure of dedicated anticoagulation clinics contributes largely to a relatively high TTR in many in the Netherlands. (34) Thus, the benefits of the DOACs as reported in international trials (in terms of the lower bleeding risk) might be limited in those with high quality of INR control. (35, 36) Plus, the INR measurements safeguard the therapeutic compliance that is important for retaining therapeutic anticoagulant levels. With the shorter half-life time of the DOACs, missing a dose of the new agents can negatively impact the quality of anticoagulation while this will not be noticed by the treating physician.

Considering these advantages and disadvantages, the actual effect of the introduction of the DOACs on health outcomes has yet to be determined. But if future research points towards the advantages of DOACs over VKA in non-controlled settings too, this might change the view on long-term anticoagulant treatment in those with the highest risks of VTE recurrence as well: when the benefits of prolonged therapy (no recurrence) are more likely to outweigh the disadvantages (bleeding events), there is more room to consider prolonged treatment.

3. Integration of prediction models for both bleeding and for recurrent events

In the final two chapters of this thesis, the emphasis was on the risk stratification of VTE recurrence, and the implications of a predicted low- or high recurrence risk on anticoagulant treatment duration. But when evaluating this risk of recurrence, it would be helpful to have a formal quantification of the individual bleeding risk as well. It is the trade-off between both risks that eventually drives the decision on treatment duration.

In atrial fibrillation, prediction models for the bleeding risk associated with the use of vitamin K antagonists have been developed (HAS-BLED (37), HEMORR2HAGES (38) and the ATRIA score (39)). Given the (expected) different subset of patients in terms of age and comorbidities, the performance of these scores in VTE patients is

unknown and should be evaluated before use in clinical practice first.

An issue in developing separate prediction models for VTE recurrence and bleeding risk is the fact that the predictors for VTE recurrence and bleeding events are largely overlapping. Predictors that increase the bleeding risk (e.g. age, vascular damage) are also associated with an increased recurrence risk. If both models indicate an increased risk of recurrent VTE and bleeding in an elderly patient with relevant comorbidity, the decision making process is more difficult.

Furthermore, it is difficult how to value the risk of one event versus another: what bleeding risk is accepted to prevent a VTE recurrence? Researchers have proposed ways to conquer this problem with a focus on complex models that both include bleeding and recurrence risks. (40) In an attempt to integrate prediction of different outcomes, this is a very first step. Analogue to the section on diagnostic prediction models in which the focus should be shifted towards a complains-centered approach, prognostic prediction modeling should be more about the combination of subjective factors like the quality of life (QoL) and objective outcomes like mortality of morbidity. Integration of these factors provides clinically relevant information for physicians and patients to base their treatment decision upon.

4. Role of GPs in VTE management

Medical specialists come across patients with VTE frequently and have vast experience in the initiation and determination of the duration of anticoagulant treatment. With only a small number of VTE cases annually, the role of general practitioner in this initial phase of VTE treatment is limited.

However, with the expectation that prolonged treatment will be opted for more often in the near future, GPs might play a role in the follow-up of these patients. While on prolonged anticoagulation, a structural reassessment of the appropriateness of the use of anticoagulation is necessary: patient characteristics are dynamic and regular evaluation of the general condition and potential newly developed co-morbidity is needed to prevent (iatrogenic) complications of inappropriate continuation of anticoagulant treatment. With the patient-context, actual medication prescriptions and insights in the patients' medical situation at one's disposal, the GP could play a pivotal role in this yearly reassessment on the long-run. Although not validated nor implemented yet, the use of a prognostic prediction model might be of value in this evaluation in future. It should be mentioned however that these models have been developed in patients 6 months after the initial VTE event, and thus the value in an annual reassessment, up to many years after the initial event, is unsure.

There is accumulating evidence on shared pathogenic mechanisms of both arterial and venous thromboembolic disease. (41, 42) Also, the risk factors for both thrombosis and bleeding in arterial and venous thromboembolism are comparable. Therefore, shared attention for both manifestations of cardiovascular disease is important. One could think of the integration of an annual reassessment of

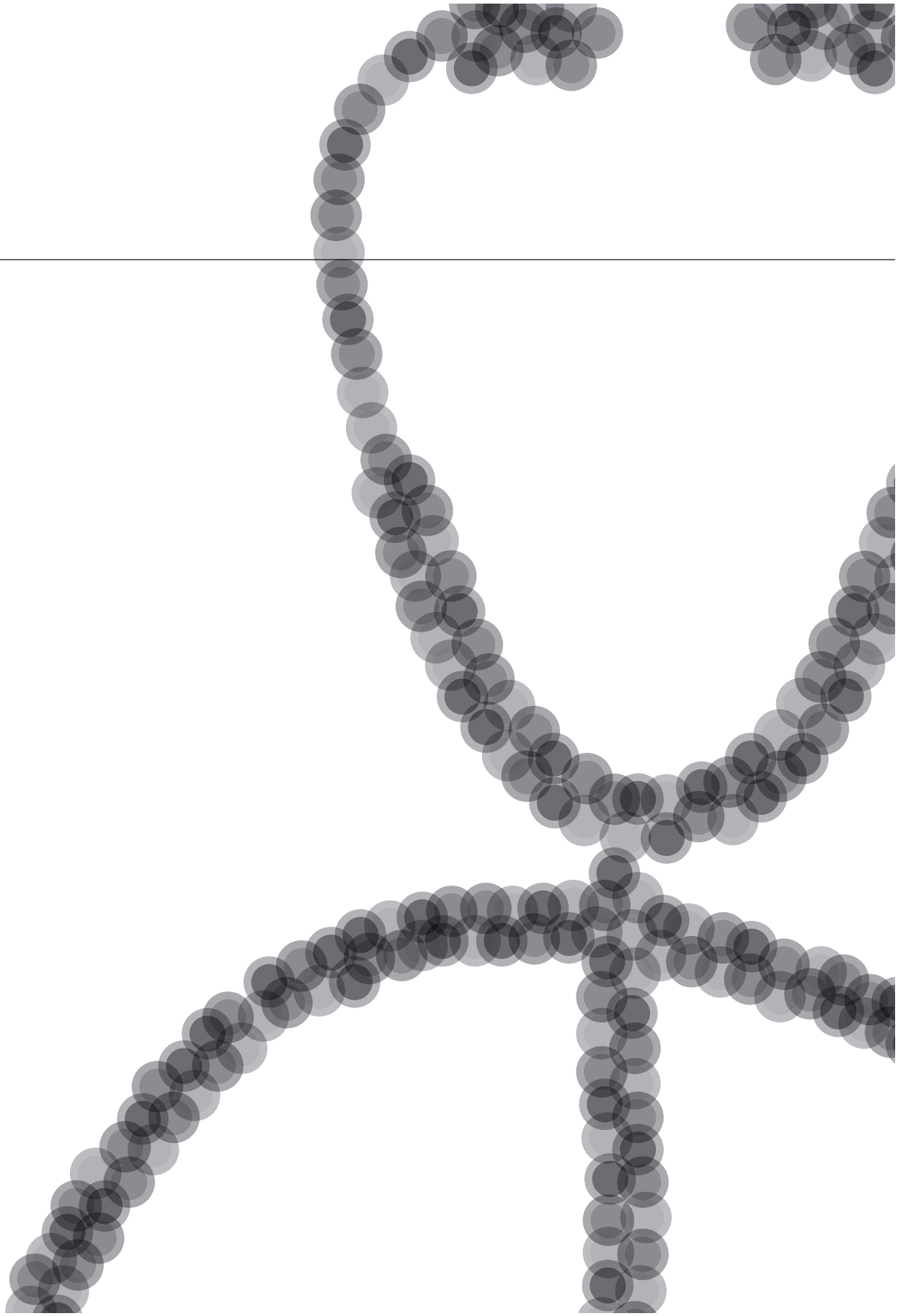
VTE patients on prolonged anticoagulation with the (already existing) yearly arterial cardiovascular disease risk evaluation in primary care. As such, GPs can structure the follow-up of VTE patients in primary care and warrant a complete cardiovascular assessment.

REFERENCES

1. Baaten G, van Bakel H, Geersing G, de Jong J, Lucassen W, van Noortwijk H, et al. Diepe veneuze trombose en longembolie (eerste herziening). Huisarts Wet. 2015;58(1):26-35.
2. Geersing GJ, Erkens PMG, Lucassen W, Buller H, Ten Cate H, Hoes A, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345:e6564.
3. Schouten HJ, Geersing GJ, Koek HL, Zuithoff NPA, Janssen KJM, Douma RA, et al. Diagnostic accuracy of conventional or age adjusted D-dimer cut-off values in older patients with suspected venous thromboembolism: systematic review and meta-analysis. BMJ. 2013;346.
4. Harper PL, Theakston E, Ahmed J, Ockelford P. D-dimer concentration increases with age reducing the clinical value of the D-dimer assay in the elderly. Intern Med J. 2007;37(9):607-13.
5. 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(6):1039-45.
6. Barritt D, Jordan S. Anticoagulant drugs in the treatment of pulmonary embolism. A controlled trial. Lancet. 1960;275(7138):1309-12.
7. Dalen J, Alpert J. Natural history of pulmonary embolism. Prog Cardiovasc Dis. 1975;17(4):259-70.
8. Douketis JD, Kearon C, Bates S, Duku EK, Ginsberg JS. Risk of fatal pulmonary embolism in patients with treated venous thromboembolism. JAMA. 1998;279(6):458-62.
9. NICE. Diagnosis and management of venous thromboembolic diseases. NICE quality standard 29. Manchester: National Institute for Health and Clinical Excellence; 2013.
10. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-based clinical practice guidelines. Chest. 2012;141(2_suppl):e419S-e94S.
11. Jiménez Castro D, Sueiro A, Díaz G, Escobar C, García-Rull S, Picher J, et al. Prognostic significance of delays in diagnosis of pulmonary embolism. Thromb Res. 2007;121(2):153-8.
12. Alonso-Martínez JL, Sánchez FJA, Echezarreta MAU. Delay and misdiagnosis in sub-massive and non-massive acute pulmonary embolism. Eur J Intern Med. 2010;21(4):278-82.
13. den Exter PL, van Es J, Erkens PMG, van Roosmalen MJG, van den Hoven P, Hovens MMC, et al. Impact of delay in clinical presentation on the diagnostic management and prognosis of patients with suspected pulmonary embolism. Am J Respir Crit Care Med. 2013;187(12):1369-73.
14. Barais M, Morio N, Cuzon Breton A, Barraine P, Calvez A, Stolper E, et al. "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(5):e98112.
15. Wells P, Andersen D, Rodger M, Ginsberg JS, Kearon C, Gent M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. Thromb Haemost. 2000;83(3):416-20.

16. Gibson NS, Sohne M, Kruip MJHA, Tick LW, Gerdes VE, Bossuyt PM, et al. Further validation and simplification of the Wells clinical decision rule in pulmonary embolism. Thromb Haemost. 2008;99:229-34.
17. Geersing GJ, Janssen KJM, Oudega R, Bax L, Hoes AW, Reitsma JB, et al. Excluding venous thromboembolism using point of care D-dimer tests in outpatients: a diagnostic meta-analysis. BMJ. 2009;339(b2880).
18. Ombandza-Moussa E, Samama M, Horellou M, Elalamy I, Conard J. Potential use of D-dimer measurement in patients treated with oral anticoagulant for a venous thromboembolic episode. Int Angiol. 2003;22(4):364-9.
19. Gremmel T, Ay C, Seidinger D, Pabinger I, Panzer S, Koppensteiner R. Soluble p-selectin, D-dimer, and high-sensitivity C-reactive protein after acute deep vein thrombosis of the lower limb. J Vasc Surg. 2011;54(6, Supplement):48S-55S.
20. Palareti G, Cosmi B, Legnani C, Tosetto A, Brusi C, Iorio A, et al. D-dimer testing to determine the duration of anticoagulation therapy. N Engl J Med. 2006;355(17):1780-9.
21. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation. 2010;121(14):1630-6.
22. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). J Thromb Haemost. 2012;10(6):1019-25.
23. Boutitie F, Pinede L, Schulman S, Agnelli G, Raskob G, Julian J, et al. Influence of preceding length of anticoagulant treatment and initial presentation of venous thromboembolism on risk of recurrence after stopping treatment: analysis of individual participants' data from seven trials. BMJ. 2011;342:d3036.
24. Palareti G, Leali N, Coccheri S, Poggi M, Manotti C, D'Angelo A, et al. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). Lancet. 1996;348(9025):423-8.
25. Landefeld C, Goldman L. Major bleeding in outpatients treated with warfarin: incidence and prediction by factors known at the start of outpatient therapy. Am J Med. 1989;87(2):144-52.
26. Pabinger I, Ay C. Biomarkers and Venous Thromboembolism. Arterioscler Thromb Vasc Biol. 2009;29(3):332-6.
27. Ay C, Jungbauer LV, Sailer T, Tengler T, Koder S, Kaider A, et al. High concentrations of soluble P-selectin are associated with risk of venous thromboembolism and the P-selectin Thr715 variant. Clin Chem. 2007;53(7):1235-43.
28. van Es N, Coppens M, Schulman S, Middeldorp S, Büller HR. Direct oral anticoagulants compared with vitamin K antagonists for acute venous thromboembolism: evidence from phase 3 trials. Blood. 2014;124(12):1968-75.
29. Agnelli G, Buller HR, Cohen A, Curto M, Gallus AS, Johnson M, et al. Oral apixaban for the teatment of acute venous thromboembolism. N Engl J Med. 2013;369(9):799-808.
30. Schulman S, Kearon C, Kakkar AK, Mismetti P, Schellong S, Eriksson H, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. N Engl J Med. 2009;361(24):2342-52.

31. EINSTEIN investigators. Oral rivaroxaban for symptomatic venous thromboembolism. N Engl J Med. 2010;363(26):2499-510.
32. EINSTEIN PE investigators. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. N Engl J Med. 2012;366(14):1287-97.
33. van der Meer FM, Rosendaal FR, Vandenbroucke JP, Briët E. Bleeding complications in oral anticoagulant therapy: An analysis of risk factors. Arch Intern Med. 1993;153(13):1557-62.
34. van Geest-Daalderop J, Sturk A, Levi M, Adriaansen H. Omvang en kwaliteit van de antistollingsbehandeling met cumarinederivaten door de Nederlandse trombosediensten. Ned Tijdschr Geneeskd 2004;148(15):730-35.
35. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control*: A systematic review and metaregression. Chest. 2006;129(5):1155-66.
36. Coppens M. New oral anticoagulants versus vitamin K antagonists in countries with good INR control. Neth J Med. 2013;71(4):168-9.
37. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation. Chest. 2010;138(5):1093-100.
38. Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, et al. Clinical classification schemes for predicting hemorrhage: Results from the National Registry of Atrial Fibrillation (NRAF). Am Heart J. 2006;151(3):713-9.
39. Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N, et al. A new risk scheme to predict warfarin-associated hemorrhage: the ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) study. J Am Coll Cardiol. 2011;58(4):395-401.
40. Kittelson JM, Spyropoulos AC, Halperin JL, Kessler CM, Schulman S, Steg G, et al. Balancing risk and benefit in venous thromboembolism trials: concept for a bivariate endpoint trial design and analytic approach. J Thromb Haemost. 2013;11(8):1443-8.
41. Lijfering WM, Flinterman LE, Vandenbroucke JP, Rosendaal FR, Cannegieter SC. Relationship between venous and arterial thrombosis: a review of the literature from a causal perspective. Semin Thromb Hemost. 2011;37(08):885-96.
42. Lowe GDO. Common risk factors for both arterial and venous thrombosis. Br J Haematol. 2008;140(5):488-95.



SUMMARY

General practitioners (GPs) play an important role in prompt diagnosis, management and counselling in venous thromboembolic disease (VTE). Accurate risk estimation is of great importance in the diagnostic and prognostic phase, as this estimated risk determines further actions: in the diagnostic phase, the main concern of a GP is to decide whether or not to refer a patient for further objective testing in secondary care, based on a global judgment of the patient's condition. Considering the often a-specific presentation of signs and symptoms of pulmonary embolism (PE), and the situational context with only limited diagnostic tools at their disposal, the diagnosis of PE is frequently a challenge in primary care. Once a diagnosis has been established, knowledge of the disease prognosis is important to be able to deal with questions on the optimal anticoagulant treatment duration or counselling. Particularly in unprovoked VTE, recurrence risk is relatively high and prolonged treatment duration might be beneficial in a subset of patients. To identify those patients who might benefit from prolonged treatment, structured risk stratification tools (e.g. a prognostic prediction model) have been developed.

In this thesis, we analyzed the diagnostic and prognostic approach in patients with suspected venous thromboembolism (VTE) in primary care, aiming to improve timely identification and to minimize over-referral.

Although acknowledged to be a diagnosis frequently missed or delayed in clinical care, limited evidence is available for the specific primary care domain. In **chapter 1**, the extent of diagnostic delay of pulmonary embolism (PE) in primary care was explored in a retrospective chart review. We identified all reported PE cases in the electronic patient records of six primary care practice based on ICPC-codings. In all cases, we reported the number of days between first presentation with complaints that could be attributed to PE and the final diagnosis, and the presence of signs, symptoms and comorbid conditions. Delay of more than 7 days between first presentation at the GP and the diagnosis of pulmonary embolism was observed in 26% of cases (33 of 128 PE cases). Age over 75 years and absence of chest complaints like pain on inspiration and chest pain were determinants associated with diagnostic delay. Also, respiratory tract infections were frequently reported prior to the diagnosis of PE. However, given the explorative design of this study, we were not able to assess causation. For GPs, awareness of the possibility of a PE as an underlying cause for complaints is important in an attempt to limit diagnostic delay, however it should be balanced with the risk of over-referral.

Chapter 2 provides a comprehensive overview of developing and validating diagnostic and prognostic models in the VTE domain. In the past 20 years, there has been much attention for risk stratification in the VTE domain: the difficulties physicians face in the diagnosis of VTE (both under- and over-diagnosis) led to the development of numerous diagnostic prediction models and the introduction of D-dimer testing. Instead of developing even more models, the focus should shift

towards efficient use of existing evidence, for example by model updating. Adequate steps in risk prediction modeling – from development to validation to impact assessment – are important to improve research in the field of thromboembolism.

Currently, multiple diagnostic prediction models to rule out PE are available. In **chapter 3**, the performance of five of these diagnostic models, in combination with point-of-care D-dimer testing, to rule out PE was evaluated for the primary care domain. The selection of evaluated prediction models was based on the degree of applicability in primary care, e.g. models without chest radiography or electrocardiography. We used the AMUSE-2 cohort, consisting of 598 patients suspected of PE in primary care by their GP, to validate the five models. The efficiency of the evaluated PE diagnostic prediction models was between 40% and 50%, implicating that 40-50% of all suspected patients could be refrained from referral to secondary care. Using one of the Wells rules (original, modified or simplified) 1.2%-1.5% of the "low-risk patients" was diagnosed with PE eventually (failure rate), , whereas the failure rate of the revised Geneva models was around 3%. A prediction model with an (expert-based) failure rate of ‹2% is often considered to be safe for use in clinical practice. As such, we suggest one of the versions of the Wells rule to be used in primary care to rule out PE.

In **chapter 4**, the use of a structured diagnostic prediction model (Wells PE rule) and the GP's 'gestalt' estimate in the diagnostic process of suspected PE were compared in the AMUSE-2 cohort. GPs estimated the probability of a PE being present on a scale from 0% to 100%, in patients with unexplained cough, dyspnea or pain on inspiration. A cut-off of ‹20% in combination with D-dimer testing was used to stratify patients into low and high predicted risk of PE and to subsequently calculate the efficiency (proportion of all patients with a low predicted PE probability) and failure rate (proportion of patients with PE despite a low predicted PE probability) of this diagnostic strategy. The Wells PE rule with cut-off ≤4, again in combination with point-of-care D-dimer testing, was used as comparator. Although both approaches are safe (failure rate 1.3% (95% CI 0.2 to 4.7%) for gestalt and 1.5% (95% CI 0.4 to 3.7%) for the Wells rule), the latter is more efficient than the GP's own gestalt estimate to rule out pulmonary embolism in suspected patients in primary care (efficiency 25% (95% CI 22 to 29%) for gestalt versus 45% (95% CI 41 to 50%) for the Wells rule).

Chapter 5 reports on the evaluation of using different types of D-dimer testing, on top of the use of a diagnostic prediction model, to rule out deep venous thrombosis (DVT) in a primary care setting. To do so, we updated a Markov model that was used in the AMUSE-1 study on the use of a diagnostic prediction model to rule out DVT in primary care. We then compared four diagnostic strategies that included point-of-care D-dimer tests (Simplify, Cardiac, Triage and Nycocard) in the GP's office with two strategies that included referral to secondary care for laboratory

based D-dimer testing. The 'Laboratory' strategy (base-case) resulted in 6.986 quality-adjusted life years at the cost of €8354 per patient. The use of point-of-care D-dimer tests resulted in comparable long-term health outcomes compared to laboratory based testing procedures. The point-of-care strategies were associated with slight cost savings, with a maximum for the Simplify strategy (–€155 (95% CI: –€246 to –€83)), and might be an alternative to referring patients to laboratory facilities in the diagnostic process of DVT.

Once the diagnosis of venous thromboembolism has been established, the patient's and GP's focus shifts towards prognosis and subsequent treatment strategies. **Chapter 6** elaborates on the concept of tailored anticoagulant treatment duration based on the assessment of risk factors for recurrent venous thromboembolism. Evident and modifiable risk factors, like malignancy, surgery or immobilization, should be assessed as the optimal treatment duration in these situations is known: long-term in case of malignancy, short-term if modifiable factors are present. If an event is unprovoked, however, the VTE recurrence risk is uncertain and the optimal treatment duration is unknown. Prognostic prediction models for recurrent VTE estimate this individual recurrence risk and might be used to tailor anticoagulant treatment duration. Prolonged duration of anticoagulation therapy can be confined to those patients with a predicted increased risk of VTE recurrence, but only after taking into account one's individual bleeding risk as well. It should be noted that these prediction models have not been validated in a Dutch setting yet, and should not be used until results are available.

In **chapter 7** we present the design of the VISTA study, a pragmatic randomized trial on the use of a prediction model on recurrent VTE to tailor anticoagulant treatment duration in patients with an unprovoked venous thromboembolic event. In the intervention group of the study, the anticoagulant treatment duration is based on the estimated 1-year VTE recurrence risk by the Vienna Prediction Model, taking into account the patient's own wishes and potential contra-indications (e.g. increased bleeding risk) for prolonged treatment. The threshold for an increased VTE recurrence risk is >5% in the first year after anticoagulant withdrawal. The control group receives care-as-usual, in which the treatment duration is based on the physician's own discretion. The primary outcome is the number of recurrent VTE events during 24 months of follow-up, and secondary outcomes include the number of bleeding events and cost-effectiveness.

Finally, the clinical implications and future perspectives of diagnostic and prognostic risk stratification of venous thromboembolism in primary care are discussed. Given the recent introduction of the Wells PE rule in the Dutch VTE guidelines for GPs, it will be important to monitor the actual implementation of the rule, and factors that hinder the use of these models in daily care. Furthermore, the moment when to suspect VTE, and thus the optimal moment when to use

a model, needs further research in order to prevent diagnostic delay. Evidence on VTE recurrence prediction is rapidly increasing, and the implications of VTE recurrence prediction for individual tailoring of anticoagulant treatment duration are currently evaluated.

NEDERLANDSE SAMENVATTING

Huisartsen spelen een belangrijke rol in de diagnose, behandeling en voorlichting van veneuze trombo-embolie (VTE). Voor zowel het stellen van de diagnose, als ook voor het bepalen van de prognose van VTE is het belangrijk om te weten hoe groot de kans is op trombose voor een individuele patiënt. Deze risico-inschatting bepaalt immers eventuele verdere actie.

Voor huisartsen is het geregeld een uitdaging om een VTE diagnose tijdig te stellen. Enerzijds omdat de klinische kenmerken van veneuze trombose, en dan met name longembolieën, regelmatig aspecifiek zijn, en anderzijds omdat huisartsen slechts een klein arsenaal van diagnostische hulpmiddelen hebben. In de diagnostische fase is een risico-inschatting van VTE daarom belangrijk om te bepalen welke patiënten verwezen moeten worden naar het ziekenhuis voor nader onderzoek.

Bij patiënten met vastgestelde VTE is kennis van de prognose belangrijk voor het geven van advies over de optimale duur van antistollingsbehandeling en voorlichting over de ziekte. Voornamelijk bij patiënten met een spontane trombose is het risico op herhaling na stoppen van de behandeling verhoogd. Voor een deel van deze groep patiënten (met het hoogste individuele risico op herhaling) kan langdurig gebruik van antistolling mogelijk gunstig zijn. Voor de identificatie van deze hoog-risico patiënten, zijn verschillende prognostische hulpmiddelen ontwikkeld. Deze hulpmiddelen worden ook wel beslisregels genoemd.

In dit proefschrift worden verschillende diagnostische en prognostische strategieën beschreven die gebruikt kunnen worden bij patiënten met een verdenking op veneuze trombose, of bij patiënten bij wie de diagnose inmiddels is vastgesteld.

Het is bekend dat de diagnose longembolie vaak wordt gemist, of vertraagd wordt vastgesteld op de spoedeisende hulp. In de huisartspraktijk is hier maar beperkt onderzoek naar gedaan. In **hoofdstuk 1** hebben wij onderzocht hoe vaak de diagnose longembolie vertraagd wordt gesteld in de eerste lijn op basis van statusonderzoek. Wij identificeerden alle gerapporteerde longemboliecasussen uit de elektronische patiëntendossiers van zes huisartspraktijken. Van alle casussen noteerden wij het aantal dagen tussen het eerste bezoek van de patiënt aan de huisarts met klachten die kunnen passen bij een longembolie en de uiteindelijke diagnose, evenals de aanwezigheid van overige klachten, symptomen en ziektes. Een vertraging van meer dan zeven dagen tussen het eerste contact met de huisarts en de diagnose longembolie werd gezien bij 26% van alle casussen (33 van 128 longembolie casussen). Bij patiënten ouder dan 75 jaar en bij patiënten zonder klachten zoals pijn bij inademen en “pijn op de borst” was er vaker sprake van diagnostische vertraging. Ook was er vaker sprake van een luchtweginfectie in de periode vóór de diagnose longembolie. Vanwege de exploratieve aard van deze studie was het niet mogelijk om te kijken naar een oorzakelijk verband.

Hoofdstuk 2 beschrijft de verschillende stappen van het ontwikkelen en valideren van diagnostische en prognostische beslisregels in het domein van VTE. In de afgelopen 20 jaar is er veel aandacht geweest voor risico-inschatting in dit specifieke domein; de moeilijkheden die artsen tegenkomen bij het stellen van de diagnose (zowel onder- als over-diagnostiek) hebben geleid tot de ontwikkeling van meerdere diagnostische beslisregels en de introductie van een bloedtest (D-dimeer bepaling). Alhoewel de neiging bestaat om steeds weer nieuwe beslisregels te ontwikkelen, is het verstandiger om de aandacht te verleggen naar efficiënt gebruik van de huidige beschikbare kennis, bijvoorbeeld door bestaande regels te updaten.

Momenteel zijn er verschillende diagnostische beslisregels beschikbaar voor het uitsluiten van de diagnose longembolie. In **hoofdstuk 3** evalueerden wij vijf van deze beslisregels voor het uitsluiten van longembolieën in de huisartspraktijk. De onderzochte beslisregels werden geselecteerd op basis van de mate van bruikbaarheid in de eerste lijn. Hierbij valt te denken aan de regels waarbij geen longfoto of hartfilmpje nodig is. Het gebruik van elke beslisregel werd volgens de richtlijnen gecombineerd met een "point-of-care" D-dimeer bloedtest. Om de vijf regels te valideren, gebruikten wij het AMUSE-2 cohort, bestaande uit 598 patiënten die door de huisarts verdacht werden van een longembolie. De efficiëntie van de geëvalueerde beslisregels voor longembolie varieerde tussen de 40-50%. Dit betekent dat 40-50% van de voor longembolie verdachte patiënten niet verwezen hoefde te worden naar het ziekenhuis voor verder onderzoek ("laag risico"). Met de 3 versies van de Wells beslisregel bleek 1,2%–1,5% van alle "laag risico" patiënten toch een longembolie te hebben. Dit is de zogenoemde "failure rate". De "failure rate" van de Geneva modellen (de 2 andere onderzochte beslisregels) lag rond de 3%. Een beslisregel met een "failure rate" kleiner dan 2% wordt door experts beschouwd als voldoende veilig voor gebruik in de praktijk. Op basis van dit geaccepteerde percentage raden wij huisartsen aan om gebruik te maken van één van de versies van de Wells regel om de diagnose longembolie uit te sluiten in de eerste lijn.

In **hoofdstuk 4** vergeleken wij het gebruik van een diagnostische beslisregel (Wells regel voor longembolie) en de klinische blik van de huisarts ("gestalt") bij voor longembolie verdachte patiënten in het AMUSE-2 cohort. Huisartsen schatten de waarschijnlijkheid van de diagnose longembolie op een schaal van 0-100% bij patiënten met onverklaarde hoest, kortademigheid of pijn bij inademing. Een afkapwaarde van ‹20%, in combinatie met een D-dimeer bloedtest, werd gebruikt om patiënten in te delen in groepen met een hoog of laag voorspeld risico op een longembolie. Vervolgens werd op basis van deze risico-inschatting de "failure rate" (proportie van alle patiënten met voorspeld laag risico op longembolie die toch een longembolie blijken te hebben) en de efficiëntie (proportie van alle patiënten met een laag voorspeld risico op longembolie) van deze diagnostische strategie berekend.

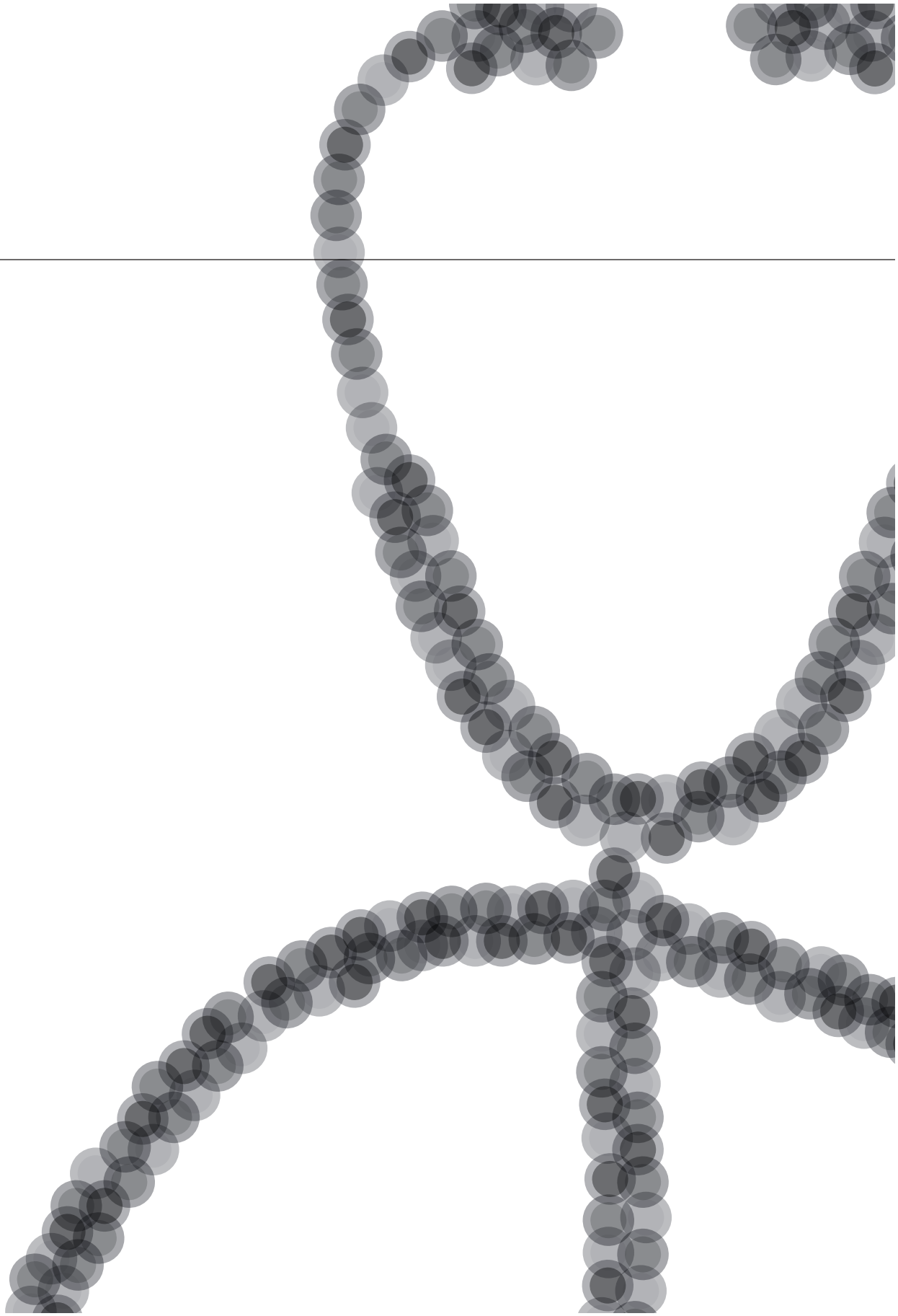
De Wells longembolie regel met afkapwaarde ≤4, gecombineerd met een D-dimeer test, werd gebruikt als vergelijkende strategie. Beide methoden waren veilig ("failure rate" 1,3% (95% betrouwbaarheidsinterval (BI) 0,2-4,7%) voor "gestalt" en 1,5% (95% BI 0,4-3,7%) voor de Wells regel), maar de Wells regel was efficiënter dan de klinische blik van de huisarts voor het uitsluiten van longembolie bij patiënten in de eerste lijn (efficiëntie 25% (95% BI 22-29%) voor "gestalt" versus 45% (95% BI 41-50%) voor de Wells regel).

Hoofdstuk 5 beschrijft de evaluatie van het gebruik van de zogenaamde D-dimeer sneltest, als aanvulling op het gebruik van een diagnostische beslisregel, voor het uitsluiten van een trombosebeen (DVT) in de huisartspraktijk. Deze sneltest kan in de spreekkamer worden gebruikt. We bewerkten een (voor de AMUSE-1 studie ontworpen) Markov model. Vervolgens vergeleken wij vier diagnostische strategieën waarbij gebruik werd gemaakt van een D-dimeer sneltest (Simplify, Cardiac, Triage en Nycocard) in de huisartspraktijk met twee strategieën waarbij patiënten werden verwezen naar het ziekenhuis voor een D-dimeer test uitgevoerd door het laboratorium. De 'Laboratorium' strategie ("base-case") resulteerde in 6.986 QALYs (een maat voor iemands kwaliteit van leven) en kostte €8354 per patiënt. Het gebruik van D-dimeer sneltesten resulteerde in vergelijkbare gezondheidsuitkomsten op de lange termijn en was geassocieerd met minimale kostenreductie, met een maximum voor de Simplify strategie (–€155 (95% BI: –€246 tot –€83)). D-dimeer sneltesten zijn een mogelijk alternatief voor het verwijzen van patiënten naar een laboratoriumfaciliteit in het diagnostische proces van DVT.

Zodra de diagnose VTE is vastgesteld, verleggen zowel de patiënt als de huisarts de focus naar de prognose en de (aan de prognose gerelateerde) behandelopties. **Hoofdstuk 6** behandelt het concept van een geïndividualiseerde duur van antistollingsbehandeling die gebaseerd is op de inschatting van risicofactoren voor een nieuwe VTE. Evidente en modificeerbare risicofactoren, zoals kanker, operatie of immobilisatie, moeten in kaart worden gebracht, omdat de optimale behandelduur in al deze gevallen duidelijk is: een langdurige behandeling in het geval van kanker, en kortdurende behandeling in het geval van de aanwezigheid van modificeerbare risicofactoren. In het geval van een niet-uitgelokte VTE is het herhalingsrisico, en daardoor ook de optimale behandelduur met antistollingsmiddelen, onzeker. Prognostische beslisregels voor herhaling van VTE schatten het herhalingsrisico voor een individuele patiënt en bieden de mogelijkheid voor een behandelduur op maat. Verlengde antistollingsbehandeling (langer dan de gebruikelijke 3-6 maanden) kan worden beperkt tot de patiënten met een voorspeld verhoogd risico op herhaling. Uiteraard moet het individuele risico op bloedingen ook worden meegenomen voordat wordt besloten tot verlengde behandeling. De tot op heden ontwikkelde beslisregels zijn nog niet gevalideerd in de Nederlandse praktijk. Daarom wordt het gebruik van prognostische beslisregels momenteel nog niet geadviseerd.

In **hoofdstuk 7** presenteren we het ontwerp van het VISTA onderzoek, een pragmatische gerandomiseerde trial naar het gebruik van het "Vienna Prediction Model" om de duur van de antistollingsbehandeling na een niet-uitgelokte VTE te bepalen. In de interventiegroep is de duur van de antistollingsbehandeling gebaseerd op het geschatte 1-jaars risico op een nieuwe VTE op basis van de beslisregel. De wens van de patiënt en potentiële contra-indicaties voor verlengde behandeling (bijvoorbeeld een verhoogd bloedingsrisico) worden meegenomen in het besluit. Als afkappunt voor een verhoogd herhalingsrisico is een risico ›5% in het eerste jaar na staken van antistollingsbehandeling gekozen. De controlegroep ontvangt gebruikelijke zorg, bestaande uit een behandelduur zoals voorgesteld door de behandelend arts op basis van zijn/ haar oordeel. Het primaire eindpunt is het aantal patiënten met een nieuwe VTE tijdens de 24 maanden dat de patiënten worden gevolgd. Verder kijken wij onder andere naar het aantal bloedingen dat optreedt en de kosteneffectiviteit.

In de algemene discussie worden de klinische implicaties en de toekomstige perspectieven van het gebruik van diagnostische en prognostische beslisregels voor VTE in de huisartspraktijk besproken. Nu de Wells regel voor longembolie sinds kort onderdeel is van de richtlijn voor huisartsen, is het belangrijk om de implementatie van deze regel in de praktijk, het effect op de patiëntenzorg en factoren die mogelijk succesvolle implementatie verhinderen, te monitoren en te evalueren. Tevens is nader onderzoek gewenst naar het optimale moment om een beslisregel in te zetten als diagnostisch hulpmiddel in de huisartspraktijk om vertraging in de diagnostiek van longembolie te voorkomen. Tot slot neemt de kennis over het voorspellen van het herhalingsrisico op VTE snel toe. Het gebruik van beslisregels voor het bepalen van een behandelduur met antistollingsmiddelen op maat wordt momenteel onderzocht.



DANKWOORD

DANKWOORD

Het is december 2015. Ik denk terug aan de bijzondere zomer van 2010. Voor een coschap was ik in Cardiff. Daar zag ik – omgeven door Spanjaarden – Oranje ten onder gaan in de WK finale voetbal. Daar voerde ik ook mijn eerste sollicitatiegesprek voor dit promotietraject. Via Skype! Vervolgens begon ik in december 2010 aan het traject. En nu is het af. Dit was nooit gelukt zonder de hulp van vele mensen.

Allereerst gaat mijn dank uit naar mijn promotieteam. Niet alleen vanwege de bereidheid om via Skype een sollicitatiegesprek te voeren, maar vooral voor de begeleiding en de mogelijkheden die ik kreeg gedurende dit traject.
Geachte prof. dr. K.G.M. Moons, beste Carl. Het is een voorrecht om bij jou te promoveren. Van een 'gewoon' paper weet jij door jouw kritische blik en handige suggesties steeds weer iets speciaals te maken. Bovendien waardeer ik jouw motiverende en opbeurende woorden als het (om welke reden dan ook) tegenzat. Dank daarvoor.
Geachte dr. R.E.G. Schutgens, beste Roger. Wat is het geweldig om jou als copromotor te hebben. Jouw enthousiasme is bijzonder aanstekelijk. Dat leidde tot een aantal mooie projecten (die helaas niet allemaal het proefschrift hebben gehaald) en congresbezoeken van Koudekerke tot Kyoto. De combinatie van sake en karaoke was hoogstwaarschijnlijk eenmalig, maar zeker onvergetelijk!
Geachte dr. G.J. Geersing, beste Geert-Jan. De hoeveelheid dank die ik je verschuldigd ben, is enorm: je was (en bleef) geduldig als ik vast dreigde te lopen in het schrijfproces, als ik niet meer geloofde in de publicatie van een artikel, of als mijn planning weer eens onhaalbaar bleek. Ik heb veel respect voor de wijze waarop jij het academisch huisartsenvak vorm geeft. Binnenkort kan niemand meer om jouw eerstelijns onderzoeksimperium heen! Veel succes daarmee.

Geachte dr. Oudega, beste Ruud. In de eerste jaren van het promotietraject werkten we nauw samen om Vista op poten te krijgen. Met jouw pragmatisme liep jij altijd tien stappen op mij voor. Dat hield de vaart erin! Alhoewel je je pensioen nog even uitstelt, wens ik je ook nu al veel ontspannen momenten toe in Nunspeet en Friesland.

Graag wil ik de leden van de beoordelingscommissie, prof. dr. G. Pasterkamp, prof. dr. F.L.J. Visseren, prof. dr. H.A.H. Kaasjager, prof. dr. N.J. de Wit en prof. dr. P.W. Kamphuisen bedanken voor de beoordeling van mijn manuscript. Ook dank ik hen en prof. dr. R.J.P. Scholten voor het zitting nemen in de oppositie tijdens de verdediging van mijn proefschrift.

Een groot deel van mijn promotietijd heb ik besteed aan Vista. Van vele kanten heb ik ondersteuning gekregen. Beste Ali, in de loop van 2011 haalde Ruud je over om je prepensioen te parkeren en voor Vista aan de slag te gaan. Een meesterzet: hij haalde een schat van kennis en ervaring in huis, maar bovenal een schat van een

vrouw. Dank je wel voor je inzet voor Vista en voor je gastvrije ontvangst (inclusief warme maaltijd) in Vierhouten.
Beste Gerry, jouw ondersteuning had op geen beter moment kunnen komen. Het was erg fijn om de logistiek met een gerust hart over te kunnen dragen.
Naast hen ben ik nog veel andere mensen dank verschuldigd: alle medewerkers van de negen deelnemende trombosediensten, in het bijzonder de medisch leiders en contactpersonen: drs. Nynke Wiersma, Vera Hagen en drs. Samia Anwar (Saltro Utrecht), dr. Ron van 't Land en Agnes Groot Bruinderink (Rode Kruis Trombosedienst Neder-Veluwe Ede), dr. Amanda Dijk, dr. Taco Bruin, Greet Bikker, Jeltje Doorn en dr. Jacoline Brinkman (Trombosedienst St. Jansdal Harderwijk), dr. Sjef van de Leur en Bea van de Schootbrugge (Isala Zwolle), dr. Rob Fijnheer en Jane Post (Meander Medisch Centrum Amersfoort), drs. Liesbeth Roos, Saffira Caffe en Alja de Ruiter (STAR-MDC Rotterdam), dr. Maarten Beinema en Ingeborg Straalman (Deventer Ziekenhuis), drs. Gert Kant, Marian Steggink, Vera Massink en Astrid Braad (MEDLON Enschede) en dr. Angelique van Holten en drs. Dineke van Dolder (Trombosedienst voor het Gooi Hilversum).
Sylvia, Ineke, Lotte, Linda en Carla voor de vele kilometers en patiëntbezoeken. Datamanagers Jildou, Yen en Joost voor de hulp bij het opzetten van de database en de bereidheid om deze database steeds verder te finetunen. Lydeke, Carla, Fien, Gerda, Truus en Dicky voor de ondersteuning en monitoring. Yvonne, Daphne, Marc en Veerle voor het invoeren van de vele vragenlijsten. De onderzoeksmedewerkers van de Van Creveldkliniek, in het bijzonder Monique Spoor, voor de goede zorgen voor de plasmasamples.
Prof. dr. F. Verheugt, prof. dr. D. Biesma en prof. dr. E. Steyerberg, voor uw tijd om zitting te nemen in de DSMB.

De bijdrage van alle coauteurs aan mijn proefschrift is aanzienlijk.
Geachte dr. Koffijberg, beste Erik. Het was een uitdaging om mijn onderzoeksperiode te starten met een kosteneffectiviteitsanalyse. Hartelijk dank voor je hulp bij alle analyses en de vele input bij het verwerken van dit alles tot een leesbaar stuk. Dr. A. Ten Cate-Hoek en dr. M. Joore, dank voor het beschikbaar stellen van het AMUSE-1 model en de suggesties voor het manuscript. Suzanne, door jouw vele werk lag er een mooi model klaar voor de uiteindelijke analyses. Dank!
Dr. Hugo Smeets en drs. Julia Velikopolskaia hielpen mij aan data uit het Julius Huisartsen Netwerk. Marc en Marleen, jullie hebben niet alleen de JHN dataset uitgeplozen, maar hebben vervolgens ook de praktijken bezocht. Dank! Deze huisartspraktijken wil ik ook bedanken voor hun medewerking.
Geachte coauteurs van de AMUSE-2 hoofdstukken, hartelijk dank voor het gebruik van de dataset en het constructief commentaar op de manuscripten.
Alhoewel het biomarkerproject het proefschrift uiteindelijk niet heeft gehaald, hebben dr. Albert Huisman, dr. Rolf Urbanus, Tesy Merkx en Fenne veel moeite gestoken in het verzamelen en analyseren van samples. Ik hoop dat we het project op een later moment alsnog kunnen afronden.

Wat zou ik moeten zonder het werk van de verschillende secretariaten. Een speciaal woord van dank voor Heinie: Je interesse en betrokkenheid, bijvoorbeeld in de vorm van een kaartje, heb ik zeer gewaardeerd. En beste Coby, altijd een glimlach en een vrolijk welkom als ik buiten adem de 6e verdieping van het Stratenum bereikte. Een reserve-Coby op de 7e van het Van Geuns mis ik nog steeds!

Tijdens de verschillende overleggen (methodologie, Geoffrey Rose en AIOTHO) binnen het Julius Centrum was er altijd ruimte voor overleg en uitwisseling van ideeën. Dank voor ieders bijdrage.

Mijn collega's dr. Peter Zuithoff, dr. Joris de Groot en (inmiddels dr.) Karlijn Groenewegen brachten licht in "R"-duisternis.

Door de inzet van Ruud en Geert-Jan kreeg ik in de afgelopen jaren steeds meer fijne VTE-collega's. Henrike, VTE-er van het eerste uur. Jij legde de lat hoog voor je opvolgers. Onze reis naar Liverpool was memorabel en ik vond het fijn om onderzoeksperikelen met je te kunnen delen. Beste Sander, geweldig om je in het VTE-team te hebben. Niet alleen vanwege je vakinhoudelijke inbreng, maar zeker ook vanwege jouw neus voor goede restaurants. Gezellig dat we de laatste maanden kamergenoten waren. Annelieke, met jou op congres gaan betekent deelnemen aan een geheel georganiseerde reis. De ISTH in Toronto was een succes; onderzoekstechnisch, maar ook appartement-met-gym-en-sauna-downtown-Toronto-technisch. Fijn dat je je ontfermt over Vista. Dank! Carline, de jongste aanwinst voor het VTE-team. Leuk om te zien dat je onderzoek vlot en goed op gang is. Veel succes met het vervolg!

Mijn (ex-)kamergenootjes Marleen, Judith, Miranda, Kim, Ellen, Joppe, Bastiaan, Marijn en de overige Toren-G-werkbuddies zorgden voor de broodnodige ontspanning tijdens al het harde werken. Het was me een waar genoeg om de werkkamer met jullie te delen. Namens "opdrinkservice van Geuns" een speciaal woord van dank voor "koffieservice van Geuns". De koffie was heerlijk.

Mijn onderzoek werd onderbroken door het eerste jaar van de huisartsopleiding in Waddinxveen. Alle lieve collega's van Medisch Centrum West zorgden ervoor dat dit jaar voorbij is gevlogen. Beste Jannie, ik waardeer je nuchtere kijk op het vak. Het was erg fijn dat je me de ruimte gaf om zelf op ontdekking te gaan. Dit gaf me veel vertrouwen. Dank!
De terugkomdagen van jaar 1 werden opgeleukt door de spontane lunches en borrels met groepje 13/6a. De congresbezoeken moeten we erin houden, inclusief etentje na afloop! De afgelopen maanden had ik veel morele steun van mijn CZ- & SEH-groepsgenoten en docenten. Op de Herstelafdeling van Zorgspectrum Nieuwegein werd mij alle ruimte geboden om de stage en proefschriftafronding te combineren. Ingrid, Jan en Lydia, hartelijk dank voor jullie begrip en de fijne samenwerking.

My Canadian family. It was awesome to meet you again in Toronto last summer.

Coenraad & Tessel, dank voor de gezellige lunches in the Basket. Ik hoop dat we deze lunchtraditie voort kunnen zetten. Sophie, de lunchpauzes in de Brink waren altijd aan de lange kant, maar ik had ze zeker niet willen missen!

Nienke G, Jef, Renze. Hoewel het soms moeite kost om de agenda's af te stemmen, hoop ik dat we de lege plekjes blijven gebruiken om af te spreken.
Gwendolyn, Else en Lothar, lief Carezza (-b, -c, enz). Het voelt vandaag als een thuiswedstrijd: geregeld hebben we galadiners in de Senaatszaal opgeluisterd met onze deuntjes. Dank jullie wel voor alle mooie muziek, maar vooral de bijzondere en vrolijke momenten daaromheen.
Anke & Ruth, onze uitstapjes naar Oerol, de dierentuin of sauna werken ontspannend en bieden de kans om bij te praten over alles wat ons bezighoudt. Dank voor jullie steun.
Het grootste deel van mijn PhD-tijd woonde ik op één van de fijnste en mooiste plekjes van Utrecht, met zes van de leukste meiden van Utrecht. Lieverds (en Keizer Karel de Grote Hopper), dank jullie wel voor alle warme jaren in een huis waar wind en kou door de kieren binnendrongen.
Alexandra, wat fijn dat je samen met Freek voor Utrecht hebt gekozen. Ik vind het geweldig om te merken dat we de draad kunnen oppakken, zelfs als we elkaar een tijdje minder zien. Dank je wel voor deze bijzondere vriendschap.

Gedeelde smart is halve smart, nietwaar lieve Judith? Onze gezamenlijke overwerk-met-pizza-avonden en schrijfweekenden hebben mij erg geholpen. Ik hoop jou ook! Veel succes volgende week.
Lieve Marleen, vanaf de 2e week van mijn onderzoek deelden wij een kamer. En deelden we alle ups-and-downs van onderzoek, onderwijs én het leven. Je bent een geweldige, attente en veel te lieve vriendin. Weet dat ik er voor je ben. Juist nu.

Mijn lieve paranimfen. Wat een geruststellende gedachte dat jullie achter mij staan.

Lieve Mariët en John, jullie enthousiasme en daadkracht zijn een welkome aanvulling aan huize Hendriksen. Ik wil jullie bedanken voor het ophangen van kapstokken en lampen, het uitzoeken van auto's (NB. volgende keer graag eentje zonder schimmel) en verhuizen, maar bovenal voor jullie steun en interesse.

Lieve Coen, grote broer! Onze squash-ochtenden mis ik, ondanks dat ik altijd compleet door je werd ingemaakt. Nu je een fiets hebt aangeschaft, kan ik niet wachten tot we samen aan de voet van de Keutenberg staan. Winst bergop compenseert alle eerdere squash-verliezen, afgesproken?!
Liefste Marieke, grote zus! Jij bent de aanstichter van het epidemiologie-avontuur door mij te wijzen op EPIC. En zo zaten we het afgelopen jaar geregeld samen

te schrijven op de zondag. Het waren waardevolle dagen voor mij: 'thesis time' gecombineerd met 'quality time'. Niettemin kijk ik vooral uit naar de zondagen waarop we gaan fietsen, Downton Abbey kijken, of met Rosalie op stap gaan. Want lieve Rosalie, het is een immens grote eer om jouw peettante te zijn!

Lieve papa, wat ben ik blij dat ik je stralend vooraan heb zien zitten bij vele concerten, mijn slagen en afstuderen. Dat beeld houd ik vandaag voor ogen. Lieve, dappere mama, wat ben ik trots op jou. De laatste jaren heb je veel voor je kiezen gekregen. Desondanks sta je altijd voor iedereen klaar met warme woorden of een lieve knuffel. Dank je wel dat je er altijd voor ons bent.

Janneke

CURRICULUM VITAE

Janneke Hendriksen was born on June 9th 1986 in Nijmegen, the Netherlands. She attended secondary school at the Stedelijk Gymnasium Nijmegen. In 2004, Janneke started her medical training at Utrecht University. During her studies, she went abroad for two internships in primary care, first in Machame, Tanzania, and then in Cardiff, Wales. In her final year of medical school, Janneke performed a research internship on the association between dairy intake and incident diabetes mellitus type 2 at the department of Cardiovascular Epidemiology of the Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht. This project was supervised by dr. ir. J.W.J. Beulens and prof. dr. Y.T. van der Schouw.

After obtaining her medical degree in December 2010, Janneke started working as a PhD student at the Julius Center. Her project on diagnostic and prognostic risk prediction of venous thromboembolism was supervised by prof. dr. K.G.M. Moons (Julius Center), dr. G.J. Geersing (Julius Center) and dr. R.E.G. Schutgens (Van Creveldkliniek, UMC Utrecht)). She combined her research project with the postgraduate MSc program Clinical Epidemiology at Utrecht University. Between September 2013 and September 2014, Janneke was enrolled in the first year of her general practitioner training. She is currently in the second year.