Excluding Deep Vein Thrombosis in primary care: Validation, updating, and implementation of a diagnostic rule

Excluding Deep Vein Thrombosis in primary care:
Validation, updating, and implementation of a diagnostic rule

Utrecht, Universiteit Utrecht, Faculteit Geneeskunde
Thesis University Utrecht - with a summary in Dutch
Proefschrift Universiteit Utrecht - met een samenvatting in het Nederlands

ISBN 978-90-393-4770-6
Author Diane Toll
Cover DMO marketing communicatie B.V., Amersfoort, the Netherlands
Lay-out Ralf Boerboom
Print Gildeprint Drukkerijen B.V., Enschede, the Netherlands

Excluding Deep Vein Thrombosis in primary care: Validation, updating, and implementation of a diagnostic rule

Uitsluiten van Diep Veneuze Trombose in de eerste lijn:
Validatie, updaten en implementatie van een diagnostische regel
(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. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 11 april 2008 des ochtends te 10.30 uur

door

Diane Barbera Toll

geboren op 15 mei 1981
te 's-Hertogenbosch

Promotoren: prof.dr. K.G.M. Moons
prof.dr. A.W. Hoes

Co-promotor: dr. R. Oudega

The research described in this thesis was funded by the Netherlands Organisation for Health Research and Development (ZonMw grant nr. 945-04-009).

Financial support by the Julius Center for Health Sciences and Primary Care for the publication of this thesis is gratefully acknowledged. Additional support was provided by Inverness Medical, GlaxoSmithKline, and Roche Nederland B.V.

Contents

Chapter 1

General Introduction

Venous thromboembolism (VTE) may manifest itself in two conditions, deep vein thrombosis (DVT) and pulmonary embolism (PE). VTE has been estimated to be the third most common vascular disorder after acute coronary syndromes and stroke,[1] affecting about 5% of the people in their liftime.[2] DVT usually starts with the formation of a thrombus occluding the deep veins of the legs, thighs, or pelvis. PE occurs if (parts of) this thrombus is dislodged from the vein wall, travels to the lungs, and lodges within the pulmonary arteries (embolisation). This can be fatal. Accurate treatment of DVT with anticoagulants can prevent thrombus progression and embolisation in the lungs.[3] However, unnecessary (anticoagulant) treatment for DVT (or PE) should be avoided because it carries the risk of (major) bleedings. Therefore, accurate and timely diagnosis of DVT is of utmost importance.[4]

In many countries, the primary care physician is the first to encounter patients with symptoms and signs suggestive of DVT. Diagnostic measurements available in primary care are limited, and the classic clinical findings from patient history and physical examination are not sufficiently accurate to diagnose or exclude DVT.[5] To limit the number of missed DVT diagnoses, physicians refer the large majority of patients -including those with a very low suspicion of DVT- for additional diagnostic

measurements, i.e. typically compression ultrasonography of the leg veins. However, more than 75% of these referred patients do not have DVT.[6–8] Hence, referring all primary care patients clinically suspected of DVT for diagnostic evaluation in secondary care is burdening for both patients and the health care system.

To improve the diagnostic process of patients suspected of having DVT, different strategies have been developed in the last decade, primarily aimed at safely excluding DVT without ultrasound testing. Given that individual symptoms and physical findings are not predictive enough to confirm or exclude the diagnosis[5], multivariable diagnostic prediction rules have been developed. These prediction rules usually combine the diagnostic information from patient history, physical examination, and D-dimer testing to estimate the probability (on a scale from 0 to 100%) that DVT is absent (or present) in an individual patient. D-dimer is a degradation product of circulating cross-linked fibrin formed during coagulation, which can be measured in individuals' blood. As D-dimer levels are typically elevated in patients with DVT, they may be helpful in ruling out the diagnosis of DVT. The so-called Wells rule -in combination with D-dimer testing- is by far the best known and most often applied strategy to safely rule out DVT in a secondary care setting.[9]

Diagnosing deep vein thrombosis in primary care patients

Diagnostic rules, however, are sensitive to differences in patient populations. Developed rules almost always perform worse when applied or tested in new patients. In particular, the performance of diagnostic rules developed in a secondary or tertiary care setting is notoriously poor in primary care because of the large differences between the clinical domains. Therefore, it is widely acknowledged that any new developed prediction rule requires formal testing of its accuracy in each setting where application is considered before implementation in daily practice.[10–15] Oudega et al.[16] tested the accuracy of the Wells rule in combination with D-dimer testing, in a large population of primary care patients suspected of DVT and found that the Wells rule does not adequately rule out DVT in primary care patients.

Because a valid diagnostic rule to adequately rule out or establish DVT in primary care was lacking, a new prediction rule was developed and internally validated (using bootstrapping techniques).[6] Independent diagnostic indicators of the presence of DVT were male gender, oral contraceptive use, presence of malignancy, recent surgery, absence of leg trauma, vein distension, calf circumference difference $\geq$ 3 cm, and abnormal D-dimer test result (Table 1.1). Application of this rule and its suggested

score threshold could reduce the number of referrals by 23%, whereas only 0.7% of the non-referred patients would have DVT.

Table 1.1: Diagnostic rule for excluding deep vein thrombosis in primary care

Predictors of deep vein thrombosis	**Weight of indicator**
Male Gender	1
Oral contraceptive use	1
Presence of malignancy	1
Recent surgery	1
Absence of leg trauma	1
Vein distension	1
Calf difference $\geq$ 3 cm	2
Abnormal D-dimer assay result	6

Score threshold: $\leq$ 3 "very low risk" at deep vein thrombosis, no referral for ultrasonosonography indicated; $\geq$ 4 "increased risk" at deep vein thrombosis, referral for ultrasonography indicated

Rationale and outline of this thesis

Diagnostic prediction rules commonly show good performance with the data from which they were developed, also after internal validation techniques have been applied to correct the rules for overfitting and overoptimism.[10–15] Good performance in the population in which the rule was developed is no guarantee that the rule will be accurate in future patients, even when applied to similar patients in the same setting. Testing the accuracy of a prediction rule in new patients, i.e. so-called external validation, is therefore necessary before implementation in daily patient care. Hence, in **chapter 2** we investigate the external validation of the primary care rule presented in Table 1.1.

Promising results on such validation studies increase the confidence that application of the diagnostic rule in daily practice may be beneficial.[13,14] It is well known, however, that the relationship between disease state and the performance of a diagnostic test or rule may change according to characteristics of the patients.[17–22] Physicians may thus rightly wonder whether the diagnostic accuracy of the primary care DVT rule varies according to clinically relevant patient characteristics. In **chapter 3**, we determine the applicability of the primary care DVT rule across clinically relevant subgroups defined by age, gender, and history of VTE.

The implementation of the primary care DVT rule in clinical practice was studied in the so-called "AMUSE study" (Amsterdam, Maastricht, Utrecht Study on the diag-

nosis of thromboEmbolism). In that study, more than 300 primary care physicians in three regions in the Netherlands were instructed to the use of the primary care DVT rule. The physicians applied the rule to over 1000 patients with signs and symptoms suggestive of DVT to determine whether referral for ultrasonography was indicated. The safety of the rule, in terms of VTE prevalence among the non-referred patients, is studied in **chapter 4**. In **chapter 5**, we tested wether simple adjustments may further improve the accuracy of the rule, i.e. reduce the proportion of missed diagnoses among the non-referred patients or increase the proportion of patients who do not need to be referred for ultrasonography. Furthermore, the AMUSE cohort is used to estimate the cost-effectiveness of the diagnostic rule for excluding DVT in primary care, as compared to usual care strategies (**chapter 6**).

As cut-off values for D-dimer tests are not standardized, we determined in **chapter 7** the optimal threshold for two frequently used laboratory D-dimer assays in diagnosing DVT. To facilitate the application of the diagnostic rule -combining patient characteristics and D-dimer testing- in the near-patient situation, several so-called "point of care" (POC) D-dimer assays have recently been introduced. POC D-dimer assays can directly be performed on whole blood with the use of hardly any additional laboratory handling. However, POC D-dimer tests can only safely substitute conventional D-dimer assays (requiring referral), if the POC assays have been shown to possess acceptable test characteristics. The diagnostic accuracy and user-friendliness of five different POC D-dimer assays is determined in a cohort of primary care patients suspected of DVT (**chapter 8**).

The general discussion consists of two parts. In **chapter 9.1**, a short overview of the consecutive phases of multivariable prediction research is provided, notably describing important aspects of validation studies and updating methods, impact analyses, and the implementation of prediction rules. In **chapter 9.2**, we briefly illustrate the consecutive phases in multivariable prediction research as described in chapter 9.1, with the diagnostic rule that has been described in chapters 1 to 6.

References

[1] National Heart L. and Institute B. Morbidity and mortality: 1998 chartbook on cardiovascular, lung and blood diseases, 1998.

[2] Spencer F. A., Emery C., Lessard D., Anderson F., Emani S., Aragam J., Becker R. C., and Goldberg R. J. The worcester venous thromboembolism study: a population-based study of the clinical epidemiology of venous thromboembolism, 2006, *J Gen Intern Med*, 21(7):722–7.

[3] Kyrle P. A. and Eichinger S. Deep vein thrombosis, 2005, *Lancet*, 365(9465):1163–74.

[4] Oudega R., Hoes A. W., Toll D. B., and Moons K. G. The value of clinical findings and d-dimer tests in diagnosing deep vein thrombosis in primary care, 2006, *Semin Thromb Hemost*, 32(7):673–7.

[5] Oudega R., Moons K. G., and Hoes A. W. Limited value of patient history and physical examination in diagnosing deep vein thrombosis in primary care, 2005, *Fam Pract*, 22(1):86–91.

[6] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[7] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[8] Wells P. S., Owen C., Doucette S., Fergusson D., and Tran H. Does this patient have deep vein thrombosis?, 2006, *Jama*, 295(2):199–207.

[9] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[10] Altman D. G. and Royston P. What do we mean by validating a prognostic model?, 2000, *Stat Med*, 19(4):453–73.

[11] Bleeker S. E., Moll H. A., Steyerberg E. W., Donders A. R., Derksen-Lubsen G., Grobbee D. E., and Moons K. G. External validation is necessary in prediction research: a clinical example, 2003, *J Clin Epidemiol*, 56(9):826–32.

[12] Justice A. C., Covinsky K. E., and Berlin J. A. Assessing the generalizability of prognostic information, 1999, *Ann Intern Med*, 130(6):515–24.

[13] McGinn T. G., Guyatt G. H., Wyer P. C., Naylor C. D., Stiell I. G., and Richardson W. S. Users' guides to the medical literature: Xxii: how to use articles about clinical decision rules. evidence-based medicine working group, 2000, *Jama*, 284(1):79–84.

[14] Reilly B. M. and Evans A. T. Translating clinical research into clinical practice: impact of using prediction rules to make decisions, 2006, *Ann Intern Med*, 144(3):201–9.

[15] Vergouwe Y., Steyerberg E. W., Eijkemans M. J., and Habbema J. D. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models, 2005, *J Clin Epidemiol*, 58(5):475–83.

[16] Oudega R., Hoes A. W., and Moons K. G. The wells rule does not adequately rule out deep venous thrombosis in primary care patients, 2005, *Ann Intern Med*, 143(2):100–7.

[17] Coughlin S. S., Trock B., Criqui M. H., Pickle L. W., Browner D., and Tefft M. C. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test, 1992, *J Clin Epidemiol*, 45(1):1–7.

[18] Detrano R., Janosi A., Lyons K. P., Marcondes G., Abbassi N., and Froelicher V. F. Factors affecting sensitivity and specificity of a diagnostic test: the exercise thallium scintigram, 1988, *Am J Med*, 84 (4):699–710.

[19] Hlatky M. A., Pryor D. B., Harrell J., Califf R. M., Mark D. B., and Rosati R. A. Factors affecting sensitivity and specificity of exercise electrocardiography. multivariable analysis, 1984, *Am J Med*, 77 (1):64–71.

[20] Levy D., Labib S. B., Anderson K. M., Christiansen J. C., Kannel W. B., and Castelli W. P. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy, 1990, *Circulation*, 81(3):815–20.

[21] Moons K. G., Es G. A., Deckers J. W., Habbema J. D., and Grobbee D. E. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example, 1997, *Epidemiology*, 8(1):12–7.

[22] Ransohoff D. F. and Feinstein A. R. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests, 1978, *N Engl J Med*, 299(17):926–30.

Part I

A rule to exclude Deep Vein Thrombosis in primary care

Chapter 2

Validation of an existing rule in suspected Deep Vein Thrombosis in primary care

The contents of this chapter are based on
Toll DB, Oudega R, Bulten RJ, Hoes AW and Moons KG
Excluding deep vein thrombosis safely in primary care
Journal of Family Practice 2006;55(7):613-8

Introduction

Most primary care patients with suspected deep vein thrombosis (DVT) -even if suspicion is low- are referred for burdensome and costly tests such as ultrasonography of the legs or venography. How many of these patients end up having DVT? About 25%.[1–3]

Clearly there is a need for a clinical tool that can help distinguish patients with low risk of DVT from those with high risk. In a previous study, we developed and internally validated a simple diagnostic prediction rule for DVT that includes 7 patient characteristics and the result of a D-dimer test (so-called derivation study).[4] In the derivation data set, this rule showed good performance and could safely exclude the presence of DVT in about one quarter of patients, minimizing the number of unnecessary patient referrals.

Prediction rules commonly show good performance with the data from which they were developed, even when bootstrapping techniques are applied to correct them for overoptimism (internal validation). But good accuracy with the development data set is no guarantee the rule will be accurate for future patients.[5–10] Testing the accuracy of a prediction rule in new patients is necessary before implementing it in daily patient care -i.e. so-called external validation or generalizability. Promising results on such validation studies increase the confidence for applying rules in practice.[8,9]

The aim of this study was to quantify (externally validate) the accuracy of our previously derived diagnostic rule for DVT in a large sample of new primary care patients suspected of having DVT, thus testing the rule's ability to safely exclude DVT.

The rule

The diagnostic rule validated in the present analysis, has been derived in a previous study by using multivariable logistic regression analysis, which identified independent diagnostic indicators of DVT: male gender, 6 items from the history and physical examination, and the result of a D-dimer test.[4] Combining these items in a prediction rule, we reached the optimal diagnostic accuracy for the diagnosis of DVT (Table 2.1). The formula of the diagnostic rule:

$$(1 \times male\ gender) + (1 \times oral\ contraceptive use) + (1 \times presence\ of\ malignancy) + (1 \times recent\ surgery) + (1 \times absence\ of\ trauma) + (1 \times vein\ distension) + (2 \times calf\ difference \geq 3\ cm) + (6 \times abnormal\ D\text{-}dimer\ test\ result)$$

The numerical value associated with each indicator represents the weight of that indicator. Each indicator is assigned the value 1 if present, and 0 if absent. For example, a man without leg trauma, with a history of malignancy, and a normal D-

dimer test result receives a score of $(1 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 0) + (2 \times 0) + (6 \times 0) = 3$ points. In the derivation study, the area under the receiver operating characteristic (ROC) curve of the rule, after adjustment for overoptimism using bootstrapping, was 0.78 (95% confidence interval [CI], 0.75 - 0.81).

To enhance the clinical usefulness of the rule, total score results were combined into different categories: very low risk (score 0-3, DVT prevalence 0.7%), low risk (score 4-6, DVT prevalence 4.5%), moderate risk (score 7-9, DVT prevalence 21.7%), and high risk (score 10-13, DVT prevalence 54.3%). Using a score threshold ≤ 3 (very low risk) and retaining these patients in primary care would result in a 23% reduction in referrals, at the cost of only 0.7% missed DVT cases. The man in the example above has a score of 3 points. The primary care physician could decide not to refer this patient, with a risk of just 0.7% of missing a DVT.

Table 2.1: Distribution of the predictors in the rule in the validation and derivation study set

Predictors of DVT	**Points attributed to predictor**	**Prevalence (%)**	
		Validation set (n=532)	**Derivation set (n=1295)** [4]
Male gender	1	40	36
Oral contraceptive use	1	9	10
Presence of malignancy	1	3	6
Recent surgery	1	13	14
Absence of leg trauma	1	83	85
Vein distension	1	18	20
Calf difference $\geq$ 3 cm	2	41	43
D-dimer abnormal*	6	75	69
DVT Presence		18	22

*D-dimer was considered abnormal if the concentration was $\geq$500 ng/ml with the Vidas assay and $\geq$400 ng/ml with the Tinaquant assay.

Patients & methods

Derivation study

The diagnostic rule was derived from 1295 consecutive patients consulting their primary care physician with symptoms suggestive of DVT. The study was performed among 110 primary care physicians affiliated with 3 non-academic hospitals. The characteristics of the derivation study and the rule are described in detail elsewhere.[4]

In short, suspicion of DVT was based on the presence of at least 1 of the following signs and symptoms: swelling, redness, or pain of the lower extremities. Patients

were included if the primary care physician decided that the diagnosis of DVT should be confirmed or excluded by objective diagnostic testing (ultrasonography) in the hospital. Exclusion criteria were symptoms or signs existing for more than 30 days or suspicion of pulmonary embolism.

From the literature, 16 history findings and physical examination items were selected as potential diagnostic indicators. After standardized history taking and physical examination, all patients were referred to the hospital to undergo D-dimer testing. Finally, repeated compression ultrasonography of the symptomatic leg was used as a reference test for all patients. D-dimer level was measured by either the ELISA method (VIDAS, bioMérieux, France) or a latex assay method (Tinaquant, Roche, Germany), depending on the lab routine of the participating hospital.

In an earlier study, the optimal thresholds of the D-dimer tests were determined; the test was considered abnormal if the latex assay yielded a D-dimer level $\geq$400 ng/ml (Tinaquant) or $\geq$500 ng/ml for the ELISA assay (VIDAS).[11] DVT was considered present if (one of) the deep veins of the legs were not or not completely compressible, as determined with a 5–7.5 MHz linear-array sonographic scanner (system VGE/Sonotion).[12] In patients with a normal ultrasound, the test was repeated within 7 days to exclude DVT.[4] All patients with a positive ultrasound were treated with anticoagulants.

Validation study

After completing the data sampling of the derivation study, we began collecting data on the next 532 consecutive patients visiting a general practitioner with symptoms suggestive of DVT. The validation study was conducted between June 1, 2003, and June 1, 2005, among the same 110 general practitioners who participated in the derivation study. The protocol of the validation study was similar to that of the original study, using the same inclusion and exclusion criteria, D-dimer assays, and definition of presence and absence of DVT. The 7 items in the rule were obtained from the standardized history taking and physical examination, and D-dimer testing and ultrasonography were performed in the hospital.

The study protocol was approved by the Medical Ethical Committee of the University Medical Centre Utrecht, and informed consent was obtained from all patients.

Statistical analysis

To quantify external validity, we calculated each patient's total score using the rule (Formula and Table 1). First, the overall ability of the rule to discriminate between

patients with and without DVT was assessed by the ROC area. Perfect discrimination is represented by an ROC area of 1.0; an ROC area of 0.5 equals the discrimination of a coin flip.[13]

All patients were assigned to 1 of 4 risk groups, based on total points scored on the rule (Table 2.2). Patients were again considered at very low risk if they received 3 points or less. For this threshold, we calculated corresponding sensitivity, specificity, negative predictive value, and likelihood ratio of a negative test result (negative likelihood ratio=[1-sensitivity]/specificity) with their 95% CIs. Because ruling out DVT is the main purpose of applying the rule, the positive predictive value and the likelihood ratio for a positive test result are not presented.

One hundred fifty-three of the 532 subjects had missing values for one or more predictors in the rule. Missing values ranged from 1.3% in gender to 11.8% for D-dimer test result (1.5% in oral contraceptive use, 2.3% trauma, 7.0% calf difference $\geq$ 3 cm, 7.5% presence of malignancy, 8.3% recent surgery, 11.7% vein distension). Data seldom are missing completely at random. Deleting subjects with a missing value not only leads to a loss of statistical power, but often also to biased results. Therefore, imputing missing values is generally preferred to complete case analysis.[14,15] Missing data were thus (single) imputed, using the linear regression method available in SPSS version 12.0.1 (SPSS, Inc, Chicago, Ill, USA). For comparison purposes, a complete case analysis was also performed.

Results

The frequencies of the diagnostic indicators in the validation (and derivation) set are presented in Table 2.1. Patients' age varied between 18 and 98 years (mean $\pm$ standard deviation: 60 $\pm$ 17 years), and 40% of the patients were male. The distribution of variables in the validation and derivation set were comparable, except for the prevalence of malignancy (3% and 6%, respectively) and DVT (18% and 22%, respectively).

Figure 2.1 displays the graph of the ROC curve of the rule. The area under the ROC curve was 0.75 (95% CI, 0.70–0.79).

In the validation population, 112 patients (21%) fell in the "very low risk" group. Application of the rule could thus reduce the number of referrals by 21% in the validation set (Table 2.2). None of these patients had DVT (Table 2.2). Accordingly, sensitivity of the rule at the threshold of $\leq$ 3 in the validation population was 100% with a corresponding specificity of 25.7%. The predictive value of a negative test result was 100%, and the negative likelihood ratio 0 (Table 2.2). With increasing rule

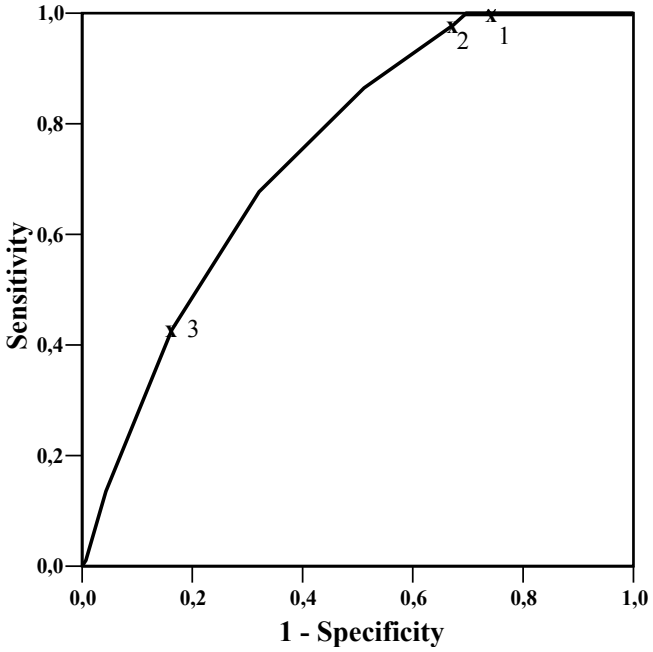

Figure 2.1: Receiver Operator Characteristic curve of the primary care rule for DVT. Points on the line correspond to the thresholds as displayed in Table 2.2. The values 1, 2 and 3 represent the sensitivities and specificities when applying a threshold of $\leq$ 3, 6 and 9, respectively.

scores, the probability of having DVT rises. Besides, if the threshold of not referring a patient would be increased, sensitivity decreases and specificity increases (Table 2.2).

The complete case analysis revealed the same sensitivity, specificity, and negative likelihood ratio as the analysis on the imputed data -i.e. no DVT cases in the very low risk group. All presented results are derived from the analysis of the imputed data set.

Discussion

To our knowledge, this diagnostic rule is the first for safely excluding DVT in primary care patients. Studies applying DVT rules in other settings have reported percentages of missed DVT cases similar to those in our study.[16–20] Of these rules, the one developed by Wells and colleagues (Wells rule) is the best known.[21–23]

The Wells rule was assembled and subsequently validated in data of 593 consecutively referred secondary care outpatients suspected of having DVT. In the Wells rule,

Table 2.2: Prevalence of DVT across the four risk categories of the rule, with accompanying diagnostic accuracy measures, in the validation and derivation study

	DVT present n(%)	DVT absent n(%)	Total number of patients (%)	Sensitivity % (95% CI)	Specificity % (95% CI)	Negative Predictive value % (95% CI)	Negative likelihood Ratio (95% CI)
Validation Study (n=532)							
Very low (0-3)	0 (0)	112 (100)	112 (21)	100 (98.6-100)	25.7 (21.6-29.8)	100 (98.8-100)	0.02 (0.00-0.32)
Low (4-6)	2 (6.3)	30 (93.8)	32 (6)	97.9 (95.1-100)	32.6 (28.1-37.0)	98.6 (96.7-100)	0.06 (0.00-0.53)
Moderate (7-9)	53 (19.2)	223 (80.8)	276 (52)	42.7 (32.8-52.6)	83.7 (80.2-87.2)	86.9 (83.7-90.1)	0.68 (0.57-0.82)
High (10-13)	41 (36.6)	71 (63.4)	112 (21)				
Derivation Study (n=1295)[4]							
Very low (0-3)	2 (0.7)	291 (0.7)	293 (23)	99.3 (98.4-100)	28.9 (26.1-31.7)	99.3 (98.4-100)	0.02 (0.00-0.10)
Low (4-6)	3 (4.5)	63 (95.5)	66 (5)				
Moderate (7-9)	144 (21.7)	519 (78.3)	663 (51)				
High (10-13)	140 (51.3)	133 (48.7)	273 (21)				

n = number of patients; 95% CI = 95% Confidence Interval

8 specific items from patient history and physical examination were weighted with 1 point (each item, if present, thus increased the likelihood of DVT equally) and a ninth item (another diagnosis just as likely or more likely to explain the presented symptoms and signs) was weighted -2 points (decreasing the likelihood of DVT). Accordingly, a sum of the scores given the 9 diagnostic items could result in a final tally from -2 to +8.

Wells et al initially presented their rule as a tool to exclude DVT if the score was 0 or lower (very low risk of DVT).[21,22] In a more recent paper they updated their rule by adding the D-dimer test result and defined patients to be at very low risk with a score of 1 or less and a normal D-dimer test result.[20] In secondary care, this rule yields good diagnostic accuracy in safely excluding DVT.

Uniqueness of our rule in primary care. In a recent study we showed that when testing the initial and updated Wells rule in primary care patients suspected of DVT, an unacceptably high percentage of patients in the "very low risk" group still had DVT: $>$ 12% for the initial rule and still $>$ 2% for the updated rule combined with D-dimer test.[24] We repeated this analysis in the current cohort of primary care patients suspected of having DVT, which yielded similar results (data not shown). The decreased accuracy of the Wells rule in primary care can probably be ascribed to the differences in spectrum of patients between secondary and primary care.[24]

Value of the validation study. Many prediction rules are developed and recommended for use in new patients without external validation. It is well known that they yield risk estimations that are too optimistic when applied to other data. Our study quantified the generalizability of our simple rule to safely exclude deep venous thrombosis in primary care.[4] Safely excluding a disease requires a high sensitivity and a high negative predictive value. The sensitivity (100%) and the negative predictive value (100%) in the new patient sample were the same as in the derivation study.

The threshold works. The derivation and validation studies both demonstrated that almost one quarter of the patients suspected of having DVT (23% and 21%, respectively) were at very low risk (score $\leq$ 3) and could safely remain in primary care (0.7% and 0% missed DVT cases, respectively). Since all 112 individuals in the validation study with a very low risk at DVT were free of DVT, sensitivity and negative predictive value of the rule in the validation study were both 100% (both 99.3% in the derivation study), with a negative likelihood ratio of 0 (table 2.2). Hence, we conclude that the developed rule seems a safe tool to use for excluding DVT in primary care.

In the validation population, only 2 patients in the "low risk" category (score $4-6$) had DVT (Table 2.2). Both patients scored 6 points on the rule. That DVT was absent in all patients with a score of $\leq$ 5 strongly confirms the safety of the chosen threshold

(i.e. not referring patients with a score of ≤ 3).

Like in the original study, we presented the distribution of patients over 4 risk categories, with accompanying diagnostic accuracy measures (Table 2.2). With increasing scores, the risk of DVT increased. When the threshold of not referring a patient would be raised, sensitivity decreases (more false negatives) and specificity increases (more true negatives). In other words, by increasing the threshold, fewer patients will be referred (saving referrals) at the cost of a higher percentage of missed DVT cases.

Referral still a judgment call. We recommend using the threshold of ≤ 3 when the rule is applied. However, use of the rule is discretionary, not mandatory. The clinical view of the physician remains important; one may still decide to refer a patient with a "very low risk" if the suspicion of DVT remains. On the other hand, circumstances may prompt a wait-and-see decision even for a patient with a score of 5.

Two caveats. First, rigorously derived prediction rules may lose their accuracy when applied to other settings, because predictors may be idiosyncratic to the population in which the rule was developed.[5,7,8] Further research could focus on whether our rule yields similar diagnostic accuracy in other settings, including secondary outpatient care.

Second, we imputed missing values in this study. Although it is acknowledged that imputing missing values is better than simply deleting all patients with one or more missing values, we repeated the entire analysis on the complete cases. The complete case analysis did not yield different results than the analysis on the imputed data.

In conclusion, this validation study demonstrated that, also in a new patient sample, the primary care physician can safely refrain from referring a considerable number of patients suspected of DVT by using a simple diagnostic rule. The use of this rule reduces the number of unnecessary patient referrals to secondary care and consequently patient burden.

References

[1] Anand S. S., Wells P. S., Hunt D., Brill-Edwards P., Cook D., and Ginsberg J. S. Does this patient have deep vein thrombosis?, 1998, *Jama*, 279(14):1094–9.

[2] Kearon C., Julian J. A., Newman T. E., and Ginsberg J. S. Noninvasive diagnosis of deep venous thrombosis. mcmaster diagnostic imaging practice guidelines initiative, 1998, *Ann Intern Med*, 128 (8):663–77.

[3] Kyrle P. A. and Eichinger S. Deep vein thrombosis, 2005, *Lancet*, 365(9465):1163–74.

[4] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[5] Altman D. G. and Royston P. What do we mean by validating a prognostic model?, 2000, *Stat Med*, 19(4):453–73.

[6] Bleeker S. E., Moll H. A., Steyerberg E. W., Donders A. R., Derksen-Lubsen G., Grobbee D. E., and Moons K. G. External validation is necessary in prediction research: a clinical example, 2003, *J Clin Epidemiol*, 56(9):826–32.

[7] Justice A. C., Covinsky K. E., and Berlin J. A. Assessing the generalizability of prognostic information, 1999, *Ann Intern Med*, 130(6):515–24.

[8] McGinn T. G., Guyatt G. H., Wyer P. C., Naylor C. D., Stiell I. G., and Richardson W. S. Users' guides to the medical literature: Xxii: how to use articles about clinical decision rules. evidence-based medicine working group, 2000, *Jama*, 284(1):79–84.

[9] Reilly B. M. and Evans A. T. Translating clinical research into clinical practice: impact of using prediction rules to make decisions, 2006, *Ann Intern Med*, 144(3):201–9.

[10] Vergouwe Y., Steyerberg E. W., Eijkemans M. J., and Habbema J. D. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models, 2005, *J Clin Epidemiol*, 58(5):475–83.

[11] Oudega R., Toll D. B., Bulten R. J., Hoes A. W., and Moons K. G. Different cut-off values for two d-dimer assays to exclude deep venous thrombosis in primary care, 2006, *Thromb Haemost*, 95(4): 744–6.

[12] Fraser J. D. and Anderson D. R. Deep venous thrombosis: recent advances and optimal investigation with us, 1999, *Radiology*, 211(1):9–24.

[13] Hanley J. A. and McNeil B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve, 1982, *Radiology*, 143(1):29–36.

[14] Greenland S. and Finkle W. D. A critical look at methods for handling missing covariates in epidemiologic regression analyses, 1995, *Am J Epidemiol*, 142(12):1255–64.

[15] Harrell J., Lee K. L., and Mark D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, 1996, *Stat Med*, 15(4): 361–87.

[16] Anderson D. R., Kovacs M. J., Kovacs G., Stiell I., Mitchell M., Khoury V., Dryer J., Ward J., and Wells P. S. Combined use of clinical assessment and d-dimer to improve the management of patients presenting to the emergency department with suspected deep vein thrombosis (the edited study), 2003, *J Thromb Haemost*, 1(4):645–51.

[17] Cornuz J., Ghali W. A., Hayoz D., Stoianov R., Depairon M., and Yersin B. Clinical prediction of deep venous thrombosis using two risk assessment methods in combination with rapid quantitative d-dimer testing, 2002, *Am J Med*, 112(3):198–203.

[18] Kraaijenhagen R. A., Piovella F., Bernardi E., Verlato F., Beckers E. A., Koopman M. M., Barone M., Camporese G., Potter Van Loon B. J., Prins M. H., Prandoni P., and Buller H. R. Simplification of the diagnostic management of suspected deep vein thrombosis, 2002, *Arch Intern Med*, 162(8):907–11.

[19] Schutgens R. E., Ackermark P., Haas F. J., Nieuwenhuis H. K., Peltenburg H. G., Pijlman A. H., Pruijm M., Oltmans R., Kelder J. C., and Biesma D. H. Combination of a normal d-dimer concentration and a non-high pretest clinical probability score is a safe strategy to exclude deep venous thrombosis, 2003, *Circulation*, 107(4):593–7.

[20] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[21] Wells P. S., Anderson D. R., Bormanis J., Guy F., Mitchell M., Gray L., Clement C., Robinson K. S., and Lewandowski B. Value of assessment of pretest probability of deep-vein thrombosis in clinical management, 1997, *Lancet*, 350(9094):1795–8.

[22] Wells P. S., Hirsh J., Anderson D. R., Lensing A. W., Foster G., Kearon C., Weitz J., D'Ovidio R., Cogo A., and Prandoni P. Accuracy of clinical assessment of deep-vein thrombosis, 1995, *Lancet*, 345 (8961):1326–30.

[23] Wells P. S., Hirsh J., Anderson D. R., Lensing A. W., Foster G., Kearon C., Weitz J., D'Ovidio R., Cogo A., Prandoni P., Girolami A., and Ginsberg J. S. A simple clinical model for the diagnosis of deep-vein thrombosis combined with impedance plethysmography: potential for an improvement in the diagnostic process, 1998, *J Intern Med*, 243(1):15–23.

[24] Oudega R., Hoes A. W., and Moons K. G. The wells rule does not adequately rule out deep venous thrombosis in primary care patients, 2005, *Ann Intern Med*, 143(2):100–7.

Chapter 3

Safely excluding Deep Vein Thrombosis in high risk groups

The contents of this chapter are based on
Toll DB, Oudega R, Vergouwe Y, Moons KGM, Hoes AW
A new diagnostic rule for deep vein thrombosis:
safety and efficiency in clinically relevant subgroups
Family Practice, accepted for publication

Introduction

The American Academy of Family Physicians and the American College of Physicians recommend that prediction rules to aid in diagnosing deep vein thrombosis (DVT) should be applied in practice.[1,2] The prediction rule developed by Wells et al.[3] (Wells rule) has been validated frequently in secondary care patients, and may therefore be the 'method of choice'. However, in primary care, this rule does not adequately rule out DVT.[4] Hence, we recently developed and externally validated a particular rule for safely excluding DVT in primary care.[5,6] This rule, which includes 7 patient history and physical examination items plus the result of a D-dimer test, discriminates 'very low' risk (non-referral) patients from patients with an increased risk of DVT. Application of this rule reduces the number of unnecessary patient referrals for ultrasound measurements and consequently patient burden, at the cost of an acceptable low proportion of DVT ($< 1\%$) in the non-referred patients (Table 3.1).

Table 3.1: Primary care DVT rule and it's performance in the derivation and validation study

Predictors of DVT	**Weight of indicator**	
Male Gender	1	
Oral contraceptive use	1	
Presence of malignancy	1	
Recent surgery	1	
Absence of leg trauma	1	
Vein distension	1	
Calf difference $\geq$ 3 cm	2	
Abnormal D-Dimer test result*	6	
SCORE		
Performance (Score cut-off: $\leq$ 3)	**Derivation Study (N = 1295)[4]**	**Validation Study (N =532)[6]**
DVT prevalence	22	18
% missed DVT	0.7	0
Proportion with low score	23	21
Sensitivity	99.3% (98.4-100)	100% (98.6-100)
Specificity	28.9 % (26.1-31.7)	25.7% (21.6-29.8)
Negative Predictive value	99.3% (98.4-100)	100% (98.8-100)

DVT = deep venous thrombosis; Score threshold: $\leq$ 3 'very low risk' at DVT, versus $\geq$ 4 'increased risk' at DVT
*D-dimer was considered abnormal if the concentration was $\geq$ 500 ng/ml with the Vidas assay and $\geq$ 400 ng/ml with the Tinaquant assay.

However, it is well known that the relationship between disease state and the performance of a diagnostic test or rule may change according to characteristics of the patients.[7–12] Physicians may thus rightly doubt whether the performance of our rule

varies when it is applied in clinically relevant patient subgroups, such as the elderly or patients with a history of DVT and/or pulmonary embolism (PE). Knowledge of these potential modifying factors enhances the implementation of the diagnostic rule.[13,14]

We had several motives to assume that there might be heterogeneity in the performance of our rule among age groups. First, the incidence of DVT increases significantly with age.[15–17] Second, compared to younger patients, elderly patients visiting the GP with signs and symptoms suggestive for DVT commonly suffer from substantial comorbidity, which can confine the symptoms of DVT. Third, the specificity of the D-dimer test -the most important predictor or test in the rule- decreases with age.[18,19] Finally, the specificity of the Wells rule for excluding DVT in secondary care has shown to decrease significantly with age.[20] In addition, we hypothesized that the performance of the rule might be different in men and women, because some risk factors for DVT are gender-specific (e.g. hormones, pregnancy).

Finally, in various studies on the diagnosis of DVT in suspected patients, patients with a history of DVT and/or PE (i.e. history of venous thromboembolism or VTE) were excluded, usually because the diagnosis in these patients is assumed to be different or perhaps more difficult.[3,21–23]

The aim of the present study was to determine the applicability of our DVT rule across clinically relevant subgroups defined by age, gender, and history of VTE. The applicability of the rule was tested by measuring the efficiency of the rule (i.e. proportion of patients designated by the rule as having a very low risk of DVT and thus do not have to be referred) and the rule's ability to safely exclude DVT (% missed DVT cases among these very low risk patients). In case of a predefined low efficiency or safety, we attempted to improve the rule by adjusting the D-dimer threshold or the rule's threshold to indicate very low risk.

Methods

We retrospectively analyzed the data of 2086 consecutive primary care patients suspected of DVT. The data originated from a large cross-sectional study on the diagnosis of DVT, executed between January 1, 2002 and January 1, 2006. Of this study, the data of the first 1295 patients were used for deriving the rule[5], and the data of subsequent 532 patients were used for externally validating the rule[6]. Hence, for present analysis, the data of the derivation and validation study were combined and supplemented with the data of 259 additional patients from the same cross-sectional study who were included after the derivation and validation study.

The characteristics of the study were described in detail elsewhere.[5,6] In short, the

study was performed in the Netherlands, among 110 primary care physicians affiliated with three non-academic hospitals. All patients clinically suspected of DVT by the primary care physician were eligible for the study. Suspicion of DVT was based on swelling, redness, or pain of the legs that had been present for not more than 30 days. Patients in whom pulmonary embolism was also suspected were excluded. Just before the start of the study, the participating primary care physicians received detailed instruction in a specially organized conference, including workshops dedicated to the logistics of the study. In addition, all GPs received similar information by mail. For motivational purposes, newsletters including information about the progress and inclusion rate of the project and feedback information were forwarded to all participating physicians and their assistants during the first year of the study.

The primary care physician systematically obtained patient history and physical examination, and subsequently referred patients to the hospital to undergo D-dimer testing. Finally, real time compression ultrasonography of the symptomatic leg was used as reference test in all patients. D-dimer level was measured by either the ELISA method (VIDAS, bioMérieux, France) or a latex assay method (Tinaquant, Roche, Germany), depending on the routine of the participating lab. In an earlier study, the optimal thresholds of the D-dimer tests were determined; the test was considered abnormal if the assay yielded a D-dimer level $\geq$ 400 ng/ml (Tinaquant assay) or $\geq$ 500 ng/ml (VIDAS).[5] DVT was considered present if (at least one of) the proximal deep veins of the legs were not or not completely compressible, as determined with a 5–7.5 MHz linear-array sonographic scanner (system V GE/Sonotion).[24] In patients with a normal ultrasound, the test was repeated within 7 days to exclude DVT.

Table 1 shows a short overview of the characteristics of the DVT rule, and the performance of the rule in the derivation and validation study. In those analyses, each patient's total rule score was dichotomized at a threshold of $\leq$ 3 to distinguish the 'very low risk' (not to be referred) patients from patients with an 'increased risk' of DVT (score $\geq$ 4; to be referred for objective diagnostics, i.e. compression ultrasonography). Both, the derivation and validation study showed that applying the rule, and retaining the very low risk patients in primary care, would result in approximately one quarter reduction in referrals, at the cost of less than 0.7% missed DVT cases.

Sample size

For the subgroup analyses presented in this study, we had access to a very large database of 2086 patients. To calculate the power to assess the safety of the DVT rule in all 7 subgroups, the percentage of missed DVT cases among the 'very low risk' patients was estimated at 0.5% (between the 0.7% found in the derivation popula-

tion[5] and 0% in the validation population[6]). To be able to exclude a percentage false negatives of 3% or higher (95% upper limit of the confidence interval), taking a type I error of 0.05 (one-sided) and a type II error of 0.2, 108 patients with a score ≤ 3 are needed in each subgroup.[25] Previous studies[5,6] showed that approximately 25% of the patients has a score ≤ 3, thus each subgroup should contain at least 432 patients to achieve enough statistical power.

Statistical analysis

We calculated for each patient the total score using the DVT rule (Table 3.1). Applying the score threshold of ≤ 3, we estimated the proportion of patients in whom DVT could be excluded according to the rule ("efficiency"), and the percentage of DVT in this non-referral group (i.e. false negatives, = 1 - negative predictive value; "safety"), with 95% confidence intervals. To study the effect of age, all patients were a priori categorised into three groups: < 50 years, 50–70 years, and > 70 years, while considering group sizes based on the a priori power considerations. Developing DVT before the age of 50 (youngest age group) is relatively rare; < 25 per 100,000 patients, while as from the age of 70 years or older (eldest age group), the incidence increases substantially to > 232 per 100,000.[15] If the age boundaries would have been chosen more extreme, some age groups would have become relatively small, compromising statistical power. Next, the analyses were repeated after the total study population was divided based on gender, and finally based on the patient's history of VTE.

It was arbitrarily decided in advance that if the proportion of patients in whom DVT can be excluded was substantially decreased ($< 15\%$) or the percentage of DVT in the non-referral group was unacceptably high ($> 3\%$), we would attempt to improve the rule's performance by adjusting either the D-dimer cut-off level or the rule's threshold for this specific subgroup.

Five hundred sixty two of the 2086 subjects had missing values for one or more predictors in the rule. Missing values ranged from 0.9% in gender to 12.2% for the D-dimer test result (2.8% in oral contraceptive use, 4.6% trauma, 6.4% calf difference ≥ 3 cm, 7.2% presence of malignancy, 7.3% recent surgery, 10.7% vein distension). Missingness of data seldom occurs completely at random. Deleting subjects with a missing value does not only lead to a loss of statistical power, but often also to biased results. Therefore, imputing missing values is generally preferred to complete case analysis.[26–29] Missing data were thus (multiple) imputed, using S-PLUS for windows, professional edition, version 6.2. Only when determining the optimal D-dimer threshold for a subgroup of patients with a low efficiency or safety, patients with missing D-dimer levels were excluded.

Results

The performance of the rule for the whole study population and the subgroups are presented in Table 3.2.

Table 3.2: Performance of the primary care DVT rule across clinically relevant subgroups

Subgroup DVT	Prevalence (95% CI)	% missed DVT* (95% CI)	Proportion very low risk (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Overall (n=2086)	19.9 (18-22)	0.9 (0-1.9)	23.1 (21-25)	99.0 (97.8-100)	28.5 (26-31)
Age					
$<$ 50 y (n=598)	16.7 (14-20)	0.6 (0-1.6)	38.1 (34-42)	98.7 (96.3-100)	45.5 (41-50)
50-70 y (n=756)	20.7 (18-24)	1.4 (0-3.2)	24.1 (21-27)	98.4 (96.2-100)	29.9 (26-34)
$\geq$ 70 y (n=732)	21.7 (19-25)	0.7 (0-3.1)	9.8 (8-12)	99.7 (98.6-100)	12.4 (10-15)
Gender					
Males (n=765)	25.4 (22-29)	0.2 (0-1.1)	16.9 (14-20)	99.9 (99.3-100)	22.6 (19-26)
Females (n=1321)	16.7 (15-19)	1.2 (0-2.5)	26.6 (24-29)	98.1 (96.1-100)	31.6 (29-35)
History of VTE					
No (n=1514)	20.1 (18-22)	0.8 (0-1.9)	22.0 (20-24)	99.1 (98.0-100)	27.4 (25-30)
Yes (n=572)	19.3 (16-23)	1.2 (0-3.5)	25.8 (22-30)	98.4 (95.2-100)	31.6 (27-36)

*DVT prevalence among patients designated by the rule as very low risk at DVT

Overall, 416 of the 2086 patients had DVT (prevalence of 19.9%). Applying the rule, DVT could be excluded in 23.1% of the patients (efficiency). Just 0.9% of these patients with a score $\leq$ 3 had DVT (safety).

The prevalence of DVT ranged from 16.7% in patients aged $<$ 50 years to 21.7% for patients of 70 years or older. The proportion of patients with a score $\leq$ 3 (efficiency) was 38.1% in the youngest age group, 24.1% in patients aged 50-70 years, and 9.8% in patients of 70 years or older. The percentage of DVT in these very low risk patients was 0.6%, 1.4%, and 0.7% respectively.

To improve the rule's efficiency in the eldest age group, we redefined the D-dimer threshold to indicate increased values to 1000 ng/ml (instead of 400 ng/ml for the tinaquant assay and 500 ng/ml for the VIDAS assay). Applying the same rule cut-

off value of ≤ 3, the proportion of patients in whom DVT could be excluded then increased from 9.8% to 33.8%, at the cost of only 1.4% missed DVT cases in this very low risk group. Further increasing the D-dimer threshold in elderly beyond 1000 ng/ml revealed an unacceptably high proportion DVT among the patients with a score ≤ 3 (Figure 3.1).

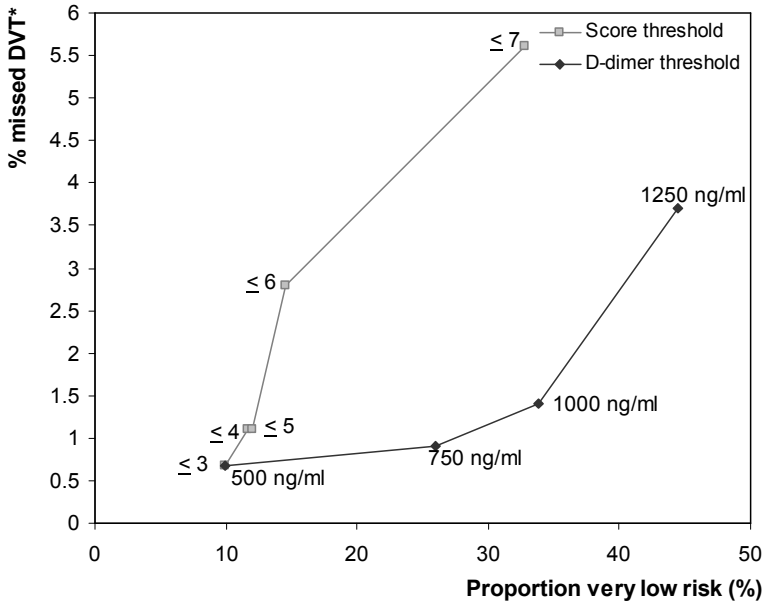

Figure 3.1: Performance of the primary care DVT rule in patients ≥ 70 years, when increasing the D-dimer or score cut-off levels. * DVT prevalence among patients designated by the rule as very low risk at DVT.

Applying a rule cut-off level of ≤ 4 or ≤ 5 instead of ≤ 3 in the elderly (without adjusting the D-dimer level) improved the efficiency of the rule only marginally. When a cut-off level of ≤ 6 was introduced, the proportion of patients in whom DVT could be excluded increased from 9.8 to 14.6%, while the percentage of missed DVT cases increased to 2.8%. When further increasing the rule cut-off level to ≤ 7, the proportion of patients in whom DVT could be excluded increased substantially to 32.8%. However, an unacceptable high proportion of these patients (5.6%) had a DVT (Figure 3.1).

The majority (63.3%) of the patients included in the study were female. The prevalence of DVT was 16.7% in females and 25.4% in males. Among the females, 26.6% had a score ≤ 3, of whom 1.2% had a DVT. Of the males, just 16.9% had a

score ≤ 3. The percentage DVT among males with a score ≤ 3 was 0.2%.

More than one quarter (27.4%) of the 2086 patients who visited the general practitioner with signs or symptoms of DVT had a history of VTE. The proportion of patients with a score ≤ 3 was lower in patients without a previous VTE (22.0%) versus patients who have had a DVT or PE before (25.8%). The percentage of patients with DVT among the non-referral patients was in both patients with and without a previous VTE around 1% (1.2% and 0.8%, respectively).

Discussion

Several diagnostic DVT rules have been developed to enhance the diagnostic work-up in patients suspected of the disease. This study quantified the performance of a recently developed simple rule to safely exclude DVT in clinically relevant subgroups of primary care patients.[4] In the literature, no formal threshold has been postulated regarding the maximum failure rate of diagnostic strategies for DVT (acceptable percentage DVT in patients in whom DVT was ruled out according to the diagnostic strategy). Preferably, the failure rate of a (new) diagnostic strategy is as low as the failure rate of the most frequently used objective diagnostic modality, i.e. serial compression ultrasonography (approximately 1%).[30,31] However, it seems reasonable that a considerable gain in efficiency of a new diagnostic strategy may counterbalance a failure rate of 2 to 3%.

In all subgroups, the failure rate was below 1.5%. This suggests that the rule can be used to safely exclude DVT in all patients, regardless of their age, gender, or history of VTE. However, the rule's efficiency did vary across subgroups. Women were more often designated by the rule as very low risk compared to men. Next, the efficiency of the rule was somewhat higher among patients with a history of VTE. Most striking was the vast decrease of the rule's efficiency with age. As the mobility of elderly is often limited, the possibility of not needing to refer a part of these patients to a hospital for further diagnostic work-up in case of suspicion of DVT is extra appealing. The decrease in efficiency could just partly be explained by the slightly higher prevalence of DVT in the eldest age group. It seemed predominantly be caused by one predictor in the rule: the D-dimer test result.

Therefore, we evaluated the consequences when a more tolerant D-dimer test threshold or score threshold was used in the eldest age group. D-dimer levels increase with age. Moreover, older patients more often have pre-existing co-morbidity, which may further elevate the D-dimer concentration.[32] Thus, elderly are more likely to have false positive D-dimer test results. We determined that in patients aged ≥ 70

years, the proportion of patients in whom DVT can be excluded can be improved from 9.8 to 33.8% without substantially compromising the safety (1.4% missed DVT cases), by raising the D-dimer threshold to 1000 ng/ml. We recommend, however, that this new D-dimer threshold in elderly should be verified prospectively to validate its safety.

Increasing the score threshold in elderly resulted in less optimistic results, either in a very small gain in efficiency, or -by increasing the score threshold even further- in too many missed DVT cases among the patients with scores below the threshold. Therefore, it is not advisable to increase the score threshold for the elderly.

The results of this study are in accordance with other studies in secondary care that evaluated the safety and efficiency of a low/moderate pretest clinical score based on the Wells rule and a negative D-dimer test result (D-dimer $<$ 500 ng/ml) in different age groups.[18,20] Both, the primary care rule[5,6] tested here and the strategy developed by Wells et al[3] can safely exclude DVT in 'very low risk' patients of all age groups in respectively primary and secondary care. Both, however, also have a significantly lower efficiency in the elderly.[20] An increased D-dimer threshold of 1000 ng/ml combined with a low/moderate Wells pretest clinical score increased the efficiency in secondary care elderly patients suspected of DVT from 12 to 28%. However, this was at the cost of 3.6% missed DVT cases.[20] The present study suggests it is safe to increase the D-dimer threshold to 1000 ng/ml in primary care elderly (gain in efficiency of 9.8 to 33.8% at the cost of 1.4% missed DVT cases). We did not find previous studies on the safety and efficiency of (other) DVT diagnostic strategies in the other studied subgroups, by gender and previous VTE.

A few methodological issues should be discussed. First, we used serial compression ultrasonography as reference test. Due to its non-invasive character and its good test characteristics, this method is widely used in clinical practice to include or exclude the diagnosis of DVT. However, it is not truly a "gold" standard, because the accuracy of serial compression ultrasonography is not 100%.[30,31] Consequently, some minor variation may be present in the 'true' performance of our rule (with all patients correctly classified as DVT present or DVT absent) compared to the results obtained in present analysis. However, it is unlikely that the safety of the rule is substantially different then presented, because the negative predictive value of serial compression ultrasonography is very high ($>$ 99%).[30,31] In no more than 1% of all patients clinically suspected of DVT, DVT is falsely excluded by serial compression ultrasonography. As the probability of DVT increases with increasing rule scores, these false negative results are more likely to occur in patients with scores above the threshold. Consequently, the true sensitivity and specificity of the rule might be slightly better than present analysis with serial compression ultrasonography as reference standard

showed. Second, it may be argued that the efficiency of the rule may be dependent on the prevalence of DVT. However, the lower the prevalence of DVT among clinically suspected patients, the more inefficient referral of all these patients for additional diagnostic testing becomes, and thus the more efficient such a rule indeed will be. Third, although our present analysis was performed on a very large cohort, validation in other primary care patient groups may consolidate the results of our study.

In conclusion, the primary care DVT rule can be used to safely exclude DVT in primary care patients suspected of the disease, regardless of age, gender and a history of VTE. Depending on the patient subgroup, up to almost 40% of the primary care patients suspected of DVT can safely be spared referral for ultrasound measurements when applying the rule.

References

[1] Qaseem A., Snow V., Barry P., Hornbake E. R., Rodnick J. E., Tobolic T., Ireland B., Segal J., Bass E., Weiss K. B., Green L., and Owens D. K. Current diagnosis of venous thromboembolism in primary care: a clinical practice guideline from the american academy of family physicians and the american college of physicians, 2007, *Ann Fam Med*, 5(1):57–62.

[2] Qaseem A., Snow V., Barry P., Hornbake E. R., Rodnick J. E., Tobolic T., Ireland B., Segal J. B., Bass E. B., Weiss K. B., Green L., and Owens D. K. Current diagnosis of venous thromboembolism in primary care: a clinical practice guideline from the american academy of family physicians and the american college of physicians, 2007, *Ann Intern Med*, 146(6):454–8.

[3] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[4] Oudega R., Hoes A. W., and Moons K. G. The wells rule does not adequately rule out deep venous thrombosis in primary care patients, 2005, *Ann Intern Med*, 143(2):100–7.

[5] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[6] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[7] Coughlin S. S., Trock B., Criqui M. H., Pickle L. W., Browner D., and Tefft M. C. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test, 1992, *J Clin Epidemiol*, 45(1):1–7.

[8] Detrano R., Janosi A., Lyons K. P., Marcondes G., Abbassi N., and Froelicher V. F. Factors affecting sensitivity and specificity of a diagnostic test: the exercise thallium scintigram, 1988, *Am J Med*, 84 (4):699–710.

[9] Hlatky M. A., Pryor D. B., Harrell J., Califf R. M., Mark D. B., and Rosati R. A. Factors affecting sensitivity and specificity of exercise electrocardiography. multivariable analysis, 1984, *Am J Med*, 77 (1):64–71.

[10] Levy D., Labib S. B., Anderson K. M., Christiansen J. C., Kannel W. B., and Castelli W. P. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy, 1990, *Circulation*, 81(3):815–20.

[11] Moons K. G., Es G. A., Deckers J. W., Habbema J. D., and Grobbee D. E. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example, 1997, *Epidemiology*, 8(1):12–7.

[12] Ransohoff D. F. and Feinstein A. R. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests, 1978, *N Engl J Med*, 299(17):926–30.

[13] Mulherin S. A. and Miller W. C. Spectrum bias or spectrum effect? subgroup variation in diagnostic test evaluation, 2002, *Ann Intern Med*, 137(7):598–602.

[14] Whiting P., Rutjes A. W., Reitsma J. B., Glas A. S., Bossuyt P. M., and Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review, 2004, *Ann Intern Med*, 140(3):189–202.

[15] Anderson J., Wheeler H. B., Goldberg R. J., Hosmer D. W., Patwardhan N. A., Jovanovic B., Forcier A., and Dalen J. E. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. the worcester DVT study, 1991, *Arch Intern Med*, 151(5):933–8.

[16] Oger E. Incidence of venous thromboembolism: a community-based study in western france. epi-getbp study group. groupe d'etude de la thrombose de bretagne occidentale, 2000, *Thromb Haemost*, 83(5):657–60.

[17] Silverstein M. D., Heit J. A., Mohr D. N., Petterson T. M., O'Fallon W. M., and Melton r. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study, 1998, *Arch Intern Med*, 158(6):585–93.

[18] Aguilar C. and Villar V. Diagnostic performance of d-dimer is lower in elderly outpatients with suspected deep venous thrombosis, 2005, *Br J Haematol*, 130(5):803–4; author reply 805.

[19] van der Graaf F., Borne H., Kolk M., Wild P. J., Janssen G. W., and Uum S. H. Exclusion of deep venous thrombosis with d-dimer testing–comparison of 13 d-dimer methods in 99 outpatients suspected of deep venous thrombosis using venography as reference standard, 2000, *Thromb Haemost*, 83(2): 191–8.

[20] Schutgens R. E., Haas F. J., and Biesma D. H. Reduced efficacy of clinical probability score and d-dimer assay in elderly subjects suspected of having deep vein thrombosis, 2005, *Br J Haematol*, 129 (5):653–7.

[21] Kahn S. R., Joseph L., Abenhaim L., and Leclerc J. R. Clinical prediction of deep vein thrombosis in patients with leg symptoms, 1999, *Thromb Haemost*, 81(3):353–7.

[22] Wells P. S., Anderson D. R., Bormanis J., Guy F., Mitchell M., Gray L., Clement C., Robinson K. S., and Lewandowski B. Value of assessment of pretest probability of deep-vein thrombosis in clinical management, 1997, *Lancet*, 350(9094):1795–8.

[23] Wells P. S., Hirsh J., Anderson D. R., Lensing A. W., Foster G., Kearon C., Weitz J., D'Ovidio R., Cogo A., and Prandoni P. Accuracy of clinical assessment of deep-vein thrombosis, 1995, *Lancet*, 345 (8961):1326–30.

[24] Fraser J. D. and Anderson D. R. Deep venous thrombosis: recent advances and optimal investigation with us, 1999, *Radiology*, 211(1):9–24.

[25] Lachin J. M. Introduction to sample size determination and power analysis for clinical trials, 1981, *Control Clin Trials*, 2(2):93–113.

[26] Donders A. R., Heijden G. J., Stijnen T., and Moons K. G. Review: a gentle introduction to imputation of missing values, 2006, *J Clin Epidemiol*, 59(10):1087–91.

[27] Greenland S. and Finkle W. D. A critical look at methods for handling missing covariates in epidemiologic regression analyses, 1995, *Am J Epidemiol*, 142(12):1255–64.

[28] Little R. Regression with missing x's: a review, 1992, *J Am Stat Assoc*, 87:1255–1264.

[29] van der Heijden G. J., Donders A. R., Stijnen T., and Moons K. G. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example, 2006, *J Clin Epidemiol*, 59(10):1102–9.

[30] Birdwell B. G., Raskob G. E., Whitsett T. L., Durica S. S., Comp P. C., George J. N., Tytle T. L., and McKee P. A. The clinical validity of normal compression ultrasonography in outpatients suspected of having deep venous thrombosis, 1998, *Ann Intern Med*, 128(1):1–7.

[31] Cogo A., Lensing A. W., Koopman M. M., Piovella F., Siragusa S., Wells P. S., Villalta S., Buller H. R., Turpie A. G., and Prandoni P. Compression ultrasonography for diagnostic management of patients with clinically suspected deep vein thrombosis: prospective cohort study, 1998, *Bmj*, 316(7124): 17–20.

[32] Sohne M., Kamphuisen P. W., Mierlo P. J., and Buller H. R. Diagnostic strategy using a modified clinical decision rule and d-dimer test to rule out pulmonary embolism in elderly in- and outpatients, 2005, *Thromb Haemost*, 94(1):206–10.

Chapter 4

Ruling out Deep Vein Thrombosis in primary care - a management study

The contents of this chapter are based on
The AMUSE study group (in alphabetical order):
Buller HR, ten Cate-Hoek AJ, Hoes AW, Joore MA,
Moons KGM, Oudega R, Prins MH, Stoffers HEJH,
Toll DB, van der Velde EF, van Weert HCPM
Safely ruling out deep venous thrombosis in primary care
Submitted

Introduction

Each year more than 140.000 inhabitants of the United Kingdom present to their primary care physician with signs and symptoms suggestive of deep venous thrombosis of the leg.[1,2] As deep venous thrombosis is a potentially life threatening disorder, current practice is to refer all these patients for objective testing to specialized diagnostic services. These services are readily available, use non-invasive tests (such as ultrasonography and D-dimer testing) and provide the referring physician with the assurance that a deep venous thrombosis is not missed.[3,4] However, numerous studies have revealed that 80 to 90% of these referred patients do not have deep venous thrombosis.[4–6] Therefore, it would be ideal to safely exclude deep venous thrombosis at initial presentation in a large proportion of these patients and thereby avoid referral.

The recent introduction of rapid point-of-care D-dimer assays combined with a specific clinical decision rule makes it possible to realize a diagnostic work-up in a primary care setting.[7–9] We, therefore, conducted a management study in a large series of consecutive patients seen in primary care to evaluate the safety and efficiency of excluding deep venous thrombosis by the combination of a clinical decision rule and a point-of-care D-dimer assay. In addition, we determined the yield of ultrasonography in the referred patients.

Methods

Setting and patients

This cohort study was conducted in primary care in the Netherlands and approximately 300 general practitioners participated. From March 2005 to January 2007, consecutive patients who presented with clinically suspected deep venous thrombosis based on the presence of at least one of the following symptoms, swelling, redness or pain of the lower extremity, were eligible for the study. Patients were excluded if they were less than 18 years of age, received anticoagulant treatment (i.e. vitamin K antagonists or low molecular weight heparin) at presentation, or were unwilling to participate. Written informed consent was obtained. The study was approved by the local review boards.

Diagnostic strategy

General practitioners applied a clinical decision rule to all study patients. This clinical decision rule was developed to safely exclude clinically suspected deep venous thrombosis in primary care patients and includes clinical items and a D-dimer assay result (Table 4.1).[9,10]

Table 4.1: Clinical Decision Rule for diagnosing deep venous thrombosis in primary care[9]

Variable	Points
Male Gender	1
Use of hormonal contraceptives	1
Active malignancy in past 6 months	1
Recent surgery in previous month	1
Absence of leg trauma	1
Distension of collateral veins	1
Difference in calf circumference $\geq$ 3 cm	2
D-dimer assay (Simplify®) abnormal	6

If score $\leq$ 3, no referral for ultrasound; If score $\geq$ 4, referral for ultrasound.

As the aim of this study was to improve the management of patients suspected of deep venous thrombosis in a primary care setting, we explicitly selected a rapid point-of-care D-dimer assay (Clearview® Simplify® D-dimer assay, Inverness Medical, Bedford, UK).[8,11] This facilitated the use of the decision rule by the general practitioner outside office hours and during house calls. A capillary blood sample was drawn by the finger prick method.[12] The test was considered abnormal if, next to the control band, a second band appeared within 10 minutes.[12] Participating physicians and their assistants received a single brief instruction on the use of the D-dimer assay and the clinical rule.

Physicians calculated for each patient the score using the clinical decision rule (Table 4.1)[10,13] and patients were managed accordingly. Those with a score $\leq$ 3 were not referred for ultrasound, received no anticoagulant treatment, but were instructed to contact their general practitioner in case of worsening symptoms. Patients with a score $\geq$ 4 were referred for ultrasound and received care as usual. Deep venous thrombosis was considered present when (one of) the proximal veins of the lower extremities was non-compressible on ultrasound.[4]

All patients were evaluated at 7 $\pm$ 2 days by the general practitioner. Three months after entering the study, all patients received a questionnaire addressing signs and symptoms of (recurrent) venous thromboembolism. Non-responders were contacted, via their general practitioners. Finally, in case of any suspicion of a (recurrent)

venous thromboembolic event during the 3 months of follow up, additional medical information of patients was retrieved from the general practitioner, including letters from hospital specialists.

Outcome measure

The primary outcome was defined as the incidence of symptomatic venous thromboembolism during 3 months of follow-up. This included fatal pulmonary embolism, nonfatal pulmonary embolism and deep venous thrombosis. An independent adjudication committee, unaware of the patient's result of the clinical decision rule, evaluated all suspected venous thromboembolic events and all deaths. A diagnosis of pulmonary embolism or deep venous thrombosis was based on a priori defined and generally accepted criteria.[14] Deaths were classified as caused by pulmonary embolism in case of confirmation by autopsy, in case of an objective test positive for pulmonary embolism prior to death, or if pulmonary embolism could not be confidently excluded as the cause of death.[14]

Statistical analysis

Based on an expected incidence of venous thromboembolism during 3 months of follow-up of 1% in patients with a score $\leq$3, and to be able to exclude a predetermined incidence of 4% or more it was calculated that 488 patients needed to be included in this low risk group (type I error 0.05, type 2 error of 0.2). The primary analysis concerned the incidence (with exact 95% confidence interval) of symptomatic venous thromboembolism during 3 months of follow-up in the group of patients with a score $\leq$ 3 who were not referred for further testing or treatment, and the percentage of patients with this score. Furthermore, we calculated the probability of venous thromboembolism on leg ultrasound at baseline or during follow up, according to the results of the clinical decision rule without the D-dimer assay result as well as to the D-dimer assay result only.

Results

Patients

A total of 1086 consecutive patients with clinically suspected deep venous thrombosis were assessed, of whom 58 (5.3%) were excluded because of predefined exclusion criteria (Figure 4.1). The characteristics of the 1028 study patients, including the

items of the clinical decision rule, are shown in Table 4.2. The mean age was 58 years and 37% were males. The suspicion of deep venous thrombosis was based most commonly on leg pain (87%) and leg swelling (78%).

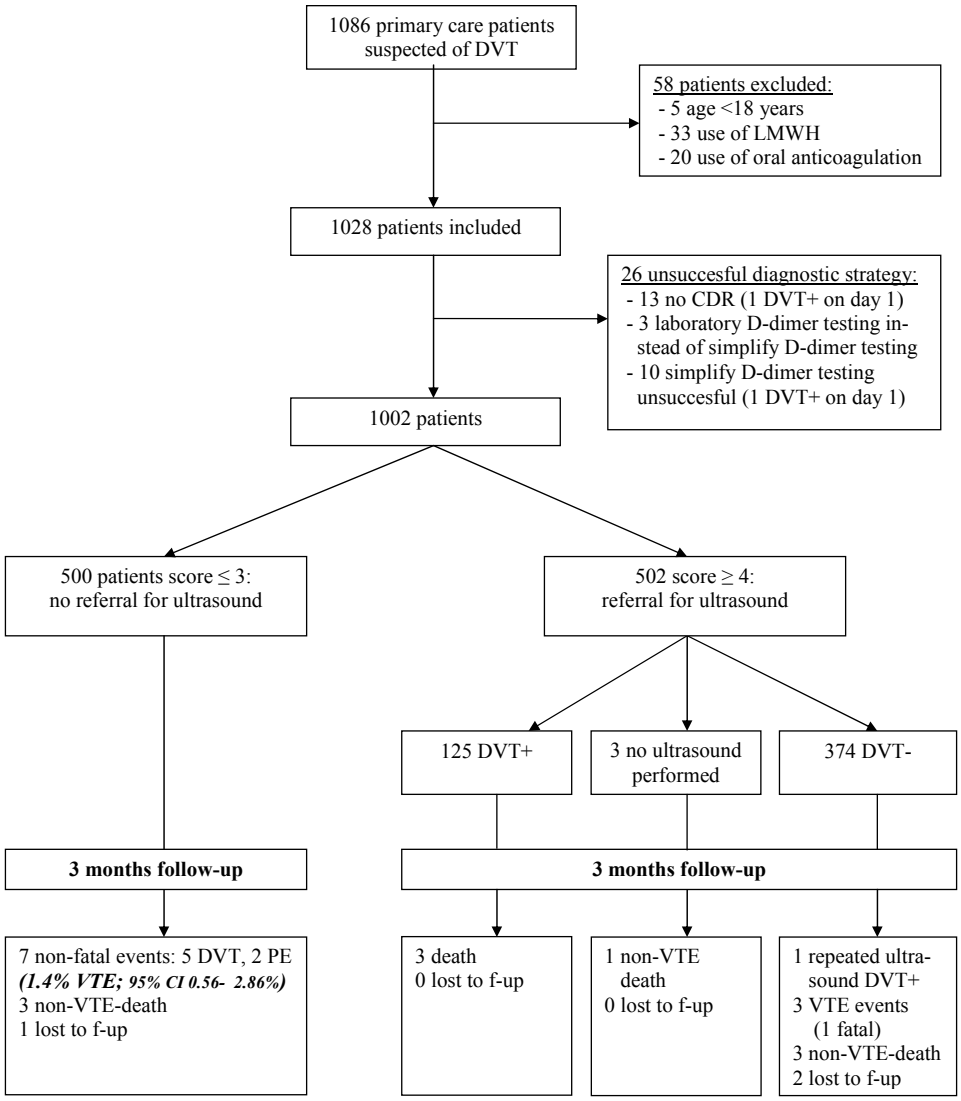

Figure 4.1: Study flowchart. Abbreviations: DVT = deep venous thrombosis; CDR = clinical decision rule; LMWH = low molecular weight heparin; DVT+ = deep venous thrombosis confirmed by ultrasonography; DVT- = deep venous thrombosis excluded by ultrasonography; VTE = venous thromboembolism; 95% CI = 95% confidence interval; f-up = follow-up

Results of Clinical Decision Rule

In 500 of the 1028 patients (49%) the score was ≤ 3 and deep venous thrombosis was considered to be ruled out (Figure 4.1). These patients were not referred for further testing and did not receive anticoagulant treatment. During the 3 month follow up, 7 of these 500 patients developed venous thromboembolism; incidence 1.4% (95% confidence interval 0.6 - 2.9%). Details of these 7 non-fatal events are given in Table 4.3.

Table 4.2: Demographic and clinical characteristics of study population (N=1028). Values are given in number (%) unless otherwise stated.

Characteristic	Value
Age, mean (SD), years	57.7 (17.1)
Male gender	375 (37%)
Leg complaints Pain of	874 (87%)
Swelling	784 (78%)
Redness	371 (37%)
Absence of leg trauma	737 (73%)
Varicose veins / venous insufficiency	337 (33%)
Distension of collateral veins	151 (15%)
Difference in calf circumference ≥ 3 cm	304 (30%)
Duration of complaints, median (IQR), days	5 (3-10)
Previous episode of DVT	159 (15%)
Previous episode of PE	51 (5%)
Paresis	13 (1%)
Recent surgery in previous month	81 (8%)
Recent immobilization	75 (7%)
Bed rest > 3 days	72 (7%)
Active malignancy in past 6 months	54 (5%)
Malignancy not treated in the past 6 months	64 (6%)
Use of hormonal contraceptives	107 (10%)
Travel (car/bus/plane) > 4 hours	90 (9%)

N = number of patients; SD = standard deviation; IQR = interquartile range

In 502 of the 1028 patients (49%) the score was ≥ 4 (Figure 4.1). An ultrasound was performed in 499 patients and showed deep venous thrombosis in 125 (25%). Of the 374 patients in whom the ultrasound was normal, 4 developed venous thromboembolism during the 3 month follow up period (1.1%, 95% confidence interval 0.3 - 2.7%).

In 26 patients (2%) the rule was not completed according to protocol (Figure 4.1). These patients were referred for ultrasound, which showed deep venous thrombosis in 2 (8%).

Table 4.3: Clinical details of the 7 patients with score ≤ 3 and a venous tromboembolic event during follow-up

Patient sex	Age, years	Score on CDR	Positive rule items	Medical history	Time to event, days	Type of event
Male	79	2	Male, absence of leg trauma	Previous TIA, use of antiplatelet drugs	2	DVT popliteal vein
Female	71	3	Vein distension, calf circumference ≥ 3 cm	Previous MI, use of antiplatelet drugs	1	DVT popliteal vein
Female	54	2	Absence of leg trauma trauma, vein distension	None	45	PE
Male	71	2	Male, absence of leg trauma	None	7	DVT popliteal vein
Male	52	2	Male, absence of leg trauma	Previous recurrent VTE	8	DVT popliteal vein
Female	44	2	Use of oral contraceptives, absence of leg trauma	None	23	DVT popliteal vein
Female	52	1	Use of oral contraceptives	None	3	PE

CDR = Clinical Decision Rule; DVT = Deep Venous Thrombosis; PE = Pulmonary Embolism; VTE = Venous Thromboembolism, TIA = Transient Ischemic Attack, MI = Myocardial Infarction;

Scenario analysis

If one would use only the clinical characteristics of the decision rule and not refer patients with a score ≤ 3, a deep venous thrombosis would be missed in 9.6% (Table 4.4). If one would use only the D-dimer assay and not refer patients with normal result, a deep venous thrombosis would be missed in 3.5%. The combination thus reduced this percentage to 1.4%. Patients with a score ≥ 4 based on clinical items only have a probability of deep venous thrombosis of 35.9%. Among these patients, those with a normal D-dimer still have a 23.5% probability of deep venous thrombosis, whereas in those with an abnormal D-dimer this probability increases to 42.6%.

Discussion

This study shows that primary care physicians can safely rule out deep venous thrombosis in approximately half of their patients by using a simple clinical decision rule

Table 4.4: Presence of venous thromboembolism either diagnosed on leg ultrasound at baseline or during the 3 months of follow up in relation to the results of the clinical decision rule without the D-dimer assay result and to the D-dimer assay results itself.

	N*	N of patients with VTE	Proportion
Clinical Items only ≤ 3	852	82	9.6%
Clinical Items only ≥ 4	145	52	35.9%
D-dimer normal	551	19	3.4%
D-dimer abnormal	446	115	25.8%
Clinical Items only ≤ 3			
D-dimer normal	500	7	1.4%
D-dimer abnormal	352	75	21.3%
Clinical Items only ≥ 4			
D-dimer normal	51	12	23.5%
D-dimer abnormal	94	40	42.6%

*In 2 of the 999 complete patients it was unknown whether the score was ≥ 4 because of the D-dimer assay or clinical items. Abbreviations: N = Number of patients; VTE = Venous Thromboembolism

combined with a point-of-care D-dimer test. The observed incidence of venous thromboembolism during three months of follow-up (1.4%; 95% C.I. 0.6-2.9) compares favourably with that in earlier studies using this strategy in referral centres, as well as with the observed incidence following normal ultrasonography in this study (1.1%, 95% C.I. 0.3-2.7).[3,6] Furthermore, our findings indicate that performing ultrasonography in the referred patients is efficient, with a confirmed thrombosis in 1 out of 4 patients. The additional benefits are that the burden on diagnostic resources will diminish and that this strategy is more convenient for patients.

Some methodological aspects of our study require comment. First, in total approximately 300 primary care physicians participated. They included a wide spectrum of consecutive patients, both during and outside office hours. The clinical characteristics of the study patients are comparable to those observed in other recent studies conducted in referral centres. Hence, we believe that our findings can be generalised to most patients suspected of deep venous thrombosis. Moreover our strategy proved to be feasible in over 97% of eligible patients (1002 of 1028 patients; Figure 4.1). The results were obtained in the setting of a true management study and participating health professionals received a single concise instruction at the start of the study only, which suggests a good prospect for implementation in daily practice.

Second, both clinical items and the point-of-care D-dimer test included in the tested strategy are important to rule out deep venous thrombosis, since the number of patients with deep venous thrombosis missed by either component alone is unac-

ceptably high (Table 4.4). Interestingly, in the presence of a high clinical score ($\geq$ 4), a normal D-dimer result is unreliable, since 24% of these patients will have deep venous thrombosis. This observation, which is in agreement with others, emphasizes that one should refrain from D-dimer testing when the clinical score is high.[6,15]

Third, we needed a D-dimer assay that uses capillary whole blood, can be performed in the general practitioners office or at a patient's home and provides an instant and easy interpretable outcome. As a result, we used a qualitative D-dimer assay.[8,16] Our results show that the Clearview® Simplify® D-dimer assay had a sensitivity of 86% and a specificity of 61% (Table 4.4). The moderate sensitivity implies that a safe exclusion of deep venous thrombosis can only be reached in patients with a low clinical score ($\leq$ 3). On the other hand the relatively high specificity results in a good clinical efficiency i.e. a large proportion of patients can be safely spared referral.

Fourth, the clinical decision rule evaluated was specifically designed for and derived in the primary care setting.[9,10] The present findings underscore the feasibility and safety of this rule.

Finally, one potential concern of the introduction of an easily accessible diagnostic test is its indiscriminate use. However, given the clinical characteristics, the overall prevalence of deep venous thrombosis (13%) and the proportion of patients with a combination of a low score on clinical items combined with a normal D-dimer test result, it is unlikely that this occurred in our study.

In conclusion, our findings indicate that primary care physicians now have a simple tool available to safely refute the diagnosis of deep venous thrombosis in a large proportion of their patients.

Acknowledgements

We gratefully acknowledge the enthusiastic participation of general practitioners and other health care professionals who contributed to the study.

This study was supported by The Netherlands Organization for Scientific Research (ZonMw), grant number 945-04-009.

References

[1] Huerta C., Johansson S., Wallander M. A., and Garcia Rodriguez L. A. Risk factors and short-term mortality of venous thromboembolism diagnosed in the primary care setting in the united kingdom, 2007, *Arch Intern Med*, 167(9):935–43.

[2] Kilroy D. A., Ireland S., Reid P., Goodacre S., and Morris F. Emergency department investigation of deep vein thrombosis, 2003, *Emerg Med J*, 20(1):29–32.

[3] Kraaijenhagen R. A., Lensing A. W., Lijmer J. G., Prandoni P., Prins M. H., Ginsberg J. S., and Buller H. R. Diagnostic strategies for the management of patients with clinically suspected deep-vein thrombosis, 1997, *Curr Opin Pulm Med*, 3(4):268–74.

[4] Lensing A. W., Prandoni P., Prins M. H., and Buller H. R. Deep-vein thrombosis, 1999, *Lancet*, 353 (9151):479–85.

[5] Ten Cate-Hoek A. J. and Prins M. H. Management studies using a combination of d-dimer test result and clinical probability to rule out venous thromboembolism: a systematic review, 2005, *J Thromb Haemost*, 3(11):2465–70.

[6] Wells P. S., Owen C., Doucette S., Fergusson D., and Tran H. Does this patient have deep vein thrombosis?, 2006, *Jama*, 295(2):199–207.

[7] Anderson D. R., Kovacs M. J., Kovacs G., Stiell I., Mitchell M., Khoury V., Dryer J., Ward J., and Wells P. S. Combined use of clinical assessment and d-dimer to improve the management of patients presenting to the emergency department with suspected deep vein thrombosis (the edited study), 2003, *J Thromb Haemost*, 1(4):645–51.

[8] Neale D., Tovey C., Vali A., Davies S., Myers K., Obiako M., Ramkumar V., and Hafiz A. Evaluation of the simplify d-dimer assay as a screening test for the diagnosis of deep vein thrombosis in an emergency department, 2004, *Emerg Med J*, 21(6):663–6.

[9] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[10] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[11] Cini M., Legnani C., Cavallaroni K., Bettini F., and Palareti G. A new rapid bedside assay for d-dimer measurement (simplify d-dimer) in the diagnostic work-up for deep vein thrombosis, 2003, *J Thromb Haemost*, 1(12):2681–3.

[12] van der Velde E. F., Wichers I. M., Toll D. B., Weert H. C., and Buller H. R. Feasibility and accuracy of a rapid 'point-of-care' d-dimer test performed with a capillary blood sample, 2007, *J Thromb Haemost*, 5(6):1327–30.

[13] Oudega R., Hoes A. W., and Moons K. G. The wells rule does not adequately rule out deep venous thrombosis in primary care patients, 2005, *Ann Intern Med*, 143(2):100–7.

[14] Buller H. R., Davidson B. L., Decousus H., Gallus A., Gent M., Piovella F., Prins M. H., Raskob G., Berg-Segers A. E., Cariou R., Leeuwenkamp O., and Lensing A. W. Subcutaneous fondaparinux versus intravenous unfractionated heparin in the initial treatment of pulmonary embolism, 2003, *N Engl J Med*, 349(18):1695–702.

[15] Parent F., Maitre S., Meyer G., Raherison C., Mal H., Lancar R., Couturaud F., Mottier D., Girard P., Simonneau G., and Leroyer C. Diagnostic value of d-dimer in patients with suspected pulmonary embolism: results from a multicentre outcome study, 2007, *Thromb Res*, 120(2):195–200.

[16] Subramaniam R. M., Heath R., Cox K., Chou T., Stewart J., and Sleigh J. Does an immunochromatographic d-dimer exclude acute lower limb deep venous thrombosis?, 2006, *Emerg Med Australas*, 18 (5-6):457–63.

Chapter 5

Optimization of the diagnostic strategy for suspected Deep Vein Thrombosis

The contents of this chapter are based on
Toll DB, van der Velde EF, ten Cate AJ, Prins MH, van Weert HCPM, Stoffers HEJH, Buller HR, Oudega R, Hoes AW, Janssen KJM, Moons KGM
Optimization of the diagnostic strategy for suspected deep vein thrombosis in primary care
Submitted

Introduction

Recently, a simple diagnostic rule or strategy was developed for general practitioners to safely exclude clinically suspected deep vein thrombosis (DVT).[1–3] The rule combines 7 clinical characteristics plus the result of a laboratory D-dimer test (Table 5.1), and was developed to discriminate patients unlikely to have DVT (who would not need additional ultrasonography) from patients with an increased risk of DVT requiring further work-up. With the recent introduction of so-called 'point of care' (POC) D-dimer tests, this diagnostic rule can be completed entirely in primary care. In a large (n = 1002) prospective management study, general practitioners actually used the diagnostic rule (Table 5.1, last column) to decide which patients suspected of DVT needed referral. The number of patient referrals for ultrasound measurements was reduced by almost 50%, at the cost of an acceptably low risk (1.4%) of subsequent venous thromboembolic events in the non-referred patients.[4] This promising result encourages the use of the diagnostic rule in clinical practice. However, it might well be that adjustment or updating of the diagnostic rule based on the new clinical data could further increase its safety or efficiency, by further reducing the proportion of missed diagnoses among the non-referred patients or increasing the proportion of patients that does not require referral for ultrasonography.

Table 5.1: The regression formula and the (simplified) scoring rule of the diagnostic algorithm developed to exclude clinically suspected deep vein thrombosis in primary care[1]

Predictors of DVT	**Regression formula** [a]	**Scoring rule** [b]
	Regression coefficient	**Weight of predictors**
Constant	-5.47	1
Male Gender	0.59	1
Oral contraceptive use	0.75	1
Presence of malignancy	0.42	1
Recent surgery	0.38	1
Absence of leg trauma	0.60	1
Vein distension	0.48	1
Calf difference $\geq$ 3 cm	1.13	2
Abnormal D-Dimer test result	3.01	6

a The regression formula of the multivariable diagnostic model (as described previously[1]) was:
Risk of DVT presence = 1/(1+exp-(-5.47 + 0.59 x male gender + 0.75 x oral contraceptive use + 0.42 x presence of malignancy + 0.38 x recent surgery + 0.60 x absence of trauma + 0.48 x vein distension + 1.13 x calf difference $\geq$ 3cm + 3.01 x abnormal D-dimer test result)).
b The weights of the predictors were based on the regression coefficients of the regression formula as described previously.[1] For each patient, a score can be calculated by counting the weights of predictors associated with the characteristics of the patient. According to the scoring rule, patients with scores $\leq$ 3 do not need to be referred for ultrasonography, in contrast to patients with scores $\geq$ 4.

This study assessed - using state of the art methodology for updating of clinical prediction rules[5–8] - whether the accuracy (notably safety and efficiency) of the diagnostic strategy could indeed be further improved or whether the original strategy represents the optimal algorithm for primary care.

Patients & Methods

Study population

We analyzed the data of the 1,002 consecutive primary care patients suspected of DVT from the above mentioned management study executed between March 2005 and January 2007, among more than 300 general practitioners in three regions in the Netherlands.[4] Patients clinically suspected of DVT, based on the presence of at least one of the following symptoms: swelling, redness or pain of the lower extremity, were eligible for the study. Patients were excluded if they were less than 18 years of age, received anticoagulant treatment (i.e. vitamin K antagonists or low molecular weight heparin) at presentation, or were unwilling to participate. Written informed consent was obtained. The study was approved by the local ethics review boards.

General practitioners systematically documented 32 variables from patient history and physical examination, including the eight predictors from the diagnostic rule and 24 other potential predictors (Table 5.2). For D-dimer measurement, a rapid POC D-dimer test (Clearview® Simplify® D-dimer, Inverness Medical, Bedford, UK),[9,10] was performed, using a capillary blood sample drawn by the fingerprick method.[11] The test was considered abnormal if, in addition to the control band, a second band appeared within 10 minutes. Participating physicians and their assistants received a brief instruction on the use of the D-dimer test and the diagnostic rule. Physicians calculated for each patient the score using the diagnostic rule (range 0-14; Table 5.1).[1,2] In accordance with the algorithm, patients with a score ≤ 3 were not referred for ultrasound and received no anticoagulant treatment, while patients with a score ≥ 4 were referred for ultrasound and received care as usual. DVT was considered present when (one of) the proximal veins of the lower extremities was non-compressible on ultrasound.[12]

All patients were evaluated one week $\pm$ 2 days after first presentation, and were followed up for 3 months to document all venous thromboembolic events. An independent adjudication committee, unaware of the patient's score of the diagnostic rule, evaluated all deaths and all suspected venous thromboembolic events. This suspicion was based on any symptom potentially indicative of VTE, reported by the patient or the physicians who were contacted by the patient. Patients with a venous throm-

boembolic event within the 3 month follow-up period were considered to have DVT at first presentation.

Methods to enhance the accuracy of the diagnostic rule

To test whether the accuracy of the previously developed algorithm (Table 5.1) could be improved, we applied so-called updating methods to the new data. Several updating methods are available.[5–8] All these methods have to and were thus applied to the underlying regression formula (Table 5.1, left column) of the diagnostic algorithm.[1] In this formula, which was obtained by using multivariable logistic regression analysis as previously described[1], the risk of DVT is defined as a function of independent predictors of DVT (Table 5.1):

$$Log((risk\ of\ DVT)/(1-risk\ of\ DVT)) = -5.47\ +\ 0.59\ \times\ male\ gender\ +\ 0.75\ \times\ oral\ contraceptive\ use\ +\ 0.42\ \times\ presence\ of\ malignancy\ +\ 0.38\ \times\ recent\ surgery\ +\ 0.60\ \times\ absence\ of\ trauma\ +0.48\ \times\ vein\ distension\ +\ 1.13\ \times\ calf\ difference\ \geq\ 3\ cm\ +\ 3.01\ \times\ abnormal\ D-dimer\ test\ result$$

The right part of this formula is called the linear predictor and includes the intercept (-5.47; to calculate the risk of DVT when all predictors are absent or zero) and the regression coefficients (relative weights or predictive strength) of the predictors.[5–8,13]

We applied three updating methods, which have all been described extensively and are here briefly explained.[5–8,13] In the first method, both the intercept and regression coefficients of the predictors were adjusted such that the calibration (i.e. the agreement between predicted risks of DVT and observed DVT frequencies) of the regression formula was optimized.[5,6] All regression coefficients were multiplied with a single correction factor that can easily be obtained from the new data (by fitting a logistic regression model with the linear predictor as the only covariate).[5,6] The intercept of the regression formula was then adjusted such that the mean predicted risk was equal to the observed DVT prevalence.[5,6] Updating method 1 will not influence the proportion of missed diagnoses among non-referred patients (safety) or the proportion of patients not requiring referral for ultrasonography (efficiency). This is because all regression coefficients of the regression formula are equally adjusted such that the relative ranking of the patients in DVT versus no DVT is not altered. Only the calibration of the rule may improve. However, this recalibration was necessary as the two following updating methods - aimed at improving the rule's safety and efficiency - build on this first updating method.

Updating method 2 was used to determine whether the strength (i.e. regression coefficient) of each of the 8 predictors in the new data was similar to the strength of the predictors in the data in which the rule was originally developed. Individual regression coefficients that were significantly different in the new data (defined by a p-value < 0.05 using the likelihood ratio test) were re-estimated or adjusted.

With updating method 3, we evaluated whether one or more of the other potential pre-

dictors (Table 5.2) that were not included in the original rule but documented in each study patient provided additional diagnostic accuracy, such that they had to be added to the original rule. For each potential predictor, the added accuracy was tested by adding each potential predictor individually to the regression formula. If the regression coefficient of a potential predictor had a p-value < 0.05 (using the likelihood ratio test) the variable was considered to significantly add to the original regression formula and was included (using forward selection starting with the potential predictor with the smallest p-value). All newly estimated regression coefficients (method 2 and 3) were finally shrunk to prevent overfitting.[5,6,14] Analyses were performed by using R version 2.5.1.

Accuracy of the updated rule

After applying the above updating methods to the underlying regression formula of the original rule, the updated regression formula was again transformed to an easy to use updated diagnostic (scoring) rule - using the same methods as applied for the original rule[1] (see also Table 5.3, legend). We then studied whether we could define a score threshold at which the updated simplified scoring rule had an improved accuracy compared to the original scoring rule of Table 5.1. Because the rule was initially developed to safely exclude suspected DVT, relevant accuracy measures of this rule were the efficiency of the rule (i.e. proportion of patients with a score below the threshold, who were thus not referred for ultrasonography) and the rule's safety (i.e. its ability to minimize the proportion of DVT cases among the patients with scores below the threshold).

Additionally, we determined the proportion of patients in whom the clinical characteristics of the final rule (without a D-dimer test result) already provided sufficient diagnostic information to determine whether patients needed to be referred for ultrasonography.

Missing data

An average of 3.3% of the values for the potential predictors was missing. Missingness of data seldom occurs completely at random. Deleting subjects with a missing value does not only lead to a loss of statistical power, but often also to biased results. Therefore, imputing missing values is generally preferred to complete case analysis.[15–18] Missing data were thus (single) imputed with values obtained from regression equations using SPSS version 14.0 (SPSS, Inc., Chicago, Il, USA).

Results

The characteristics of the 1,002 study patients, including the items from the original rule and all potential additional predictors, are shown in Table 5.2. The mean age was 57 years and 63% was female. DVT was present in 136 (13.6%) patients.

Table 5.2: Patient characteristics of the study population (N=1,002). Values are given in percentages, unless otherwise stated.

Characteristic	Value
Predictors of the diagnostic rule	
Male gender	37
Oral contraceptive use (% of total women)	17
Active malignancy in past 6 months	5
Recent surgery in previous month	8
Absence of leg trauma	73
Vein distension	15
Calf difference $\geq$ 3 cm	30
D-dimer test result abnormal	45
Potential additional predictors	
Mean age, years (SD)	57 (17)
Leg complaints Pain	87
Swelling	78
Redness	37
Acutely developed	54
Side of the affected limb, right	49
Median duration of symptoms, days (IQR)	5 (3-10)
History of DVT	16
History of PE	5
History of thrombophlebitis	8
History of any malignancy	6
Immobilization during the previous month	8
Paresis of the leg(s) during the previous month	1
Bedridden during the previous month	7
Known with hereditary coagulation disorder	2
Prolonged travelling ($>$ 4 hours seated)	9
Use of anti-platelet drugs	16
Use of NSAID	12
Pain on palpation	55
Swelling of the whole affected limb	22
Oedema	68
Varicose veins	35
Use of hormonal substitution therapy	4
Pregnancy or post-partum	2
Outcome	
Deep vein thrombosis	14

Updating of the diagnostic rule

None of the eight individual predictors of the original rule (Table 5.1) showed a significantly different effect in the new data compared to the data in which the rule was developed (updating method 2). P-values of the likelihood ratio tests varied from 0.73 for the predictor 'vein distention' to 0.12 for 'abnormal D-dimer test result'.

Updating method 3 yielded that two of the other potential predictors (Table 5.2) had significant added diagnostic value in the new data ('history of DVT', p-value 0.014 and 'prolonged travelling', p-value 0.023). Hence, these two predictors were added to the original rule. This new updated regression formula was then transformed to an updated scoring rule (Table 5.3). Because two predictors were added, the possible score range of the updated scoring rule was

Table 5.3: Regression coefficients of the original and updated regression formula and weights of the original and updated simplified scoring rule.

Predictors of DVT	**Original**		**Updated**	
	Formula, regression coefficient	Diagnostic rule, weight of predictor	Formula, regression coefficient	Diagnostic rule, weight of predictor[a]
Constant	-5.47	-	-4.58	-
Male Gender	0.59	1	0.44	1
Oral contraceptive use	0.75	1	0.56	1
Presence of malignancy	0.42	1	0.32	1
Recent surgery	0.38	1	0.29	1
Absence of leg trauma	0.60	1	0.45	1
Vein distension	0.48	1	0.36	1
Calf difference $\geq$ 3 cm	1.13	2	0.85	2
Abnormal D-Dimer test result	3.01	6	2.27	6
Prolonged travelling ($>$ 4 hours seated)	-	-	0.73	2
History of DVT	-	-	0.64	2

[a] The weight of the predictors were obtained by dividing the regression coefficient by 0.4. These figures were subsequently rounded to the nearest integer.

wider (0-18) compared to the original diagnostic rule (0-14). Applying a score of ≤ 3 to decide upon further work-up, the updated diagnostic rule excluded DVT in 43.5% of the patients (efficiency), of whom 1.4% had DVT (safety), compared to 49.4% and 1.4% by the original scoring rule, respectively (Table 5.4). After increasing the threshold of the updated rule to 4 (i.e. non-referral for scores ≤ 4), the rule's efficiency increased to 51%, at the cost of a safety of 2.9%. As expected, safety worsened further if higher thresholds for non-referral were applied to the updated scoring rule. Also when using a threshold for non-referral below 3, the updated diagnostic rule did not yield better results than the original diagnostic rule (Table 5.4).

Final diagnostic strategy

In view of the findings presented above and given the focused use of the diagnostic rule, i.e. to safely and as efficiently as possible exclude clinically suspected DVT in primary care, we decided that the original rule required no further updating. A disadvantage of the original rule is that in each patient a (POC) D-dimer test had to be executed. Table 5.5 shows that in

Table 5.4: Safety and efficiency of the original (Table 5.1) and updated (Table 5.3) diagnostic scoring rule. Values are given in percentages.

Score threshold	**Original Diagnostic scoring rule**			**Updated diagnostic scoring rule**		
	Efficiency[a]	Safety[b]	n DVT/n below threshold	Efficiency[a]	Safety[b]	n DVT/n below threshold
≤ 1	24.2	0.4	1 / 242	18.1	0.6	1 / 181
≤ 2	42.0	1.4	6 / 421	33.4	1.5	5 / 335
≤ 3	49.4	1.4	7 / 495	43.5	1.4	6 / 436
≤ 4	53.2	3.4	18 / 533	51.0	2.9	15 / 511
≤ 5	54.6	3.7	20 / 547	53.3	3.0	16 / 534
≤ 6	57.2	4.0	23 / 573	56.0	3.9	22 / 561
≤ 7	67.1	5.8	39 / 672	63.8	5.0	32 / 639
≤ 8	90.2	10.5	95 / 904	84.2	8.1	68 / 844
≤ 10	97.4	12.5	122 / 976	93.3	11.1	104 / 935
≤ 11	99.6	13.4	134 / 998	97.0	12.6	122 / 972
≤ 12	100	13.6	136 / 1002	98.9	13.2	131 / 991
≤ 13				99.9	13.6	136 / 1001
≤ 14				99.9	13.6	136 / 1001
≤ 15				100	13.6	136 / 1002

[a] Efficiency: proportion of patients with a score below the threshold of the scoring rule (non referral patients)

[b] Safety: proportion DVT among non referral patients

patients with a score of 4 points or more based on the clinical characteristics only, the risk of DVT was 35.3%, which decreased to 23.6% with a normal D-dimer test result and increased to 42.1% with an abnormal D-dimer test result. Thus, D-dimer testing is of limited value in patients with a clinical score (based on clinical characteristics only) exceeding 3 and could be restricted to patients with a clinical score ≤ 3. Figure 5.1 reflects this suggested strategy for primary care practice.

Discussion

We studied whether the accuracy of a recently developed diagnostic rule to safely exclude clinically suspected DVT could be further improved by adjusting the weights of the included predictors or by adding new diagnostic predictors. The weights of the eight individual predictors did not need to be adjusted, but inclusion of 'history of DVT' and 'prolonged travelling' significantly added predictive value. Although these new diagnostic predictors were statistically contributing to the prediction of the presence or absence of DVT and may improve the accuracy of the predicted risks on a continuous scale (0-100% risk of DVT), adding these to the rule did not confer clinically relevant effects. After introducing a threshold used in practice to determine which patients need to be referred to secondary care, the safety and efficiency of the updated rule did not improve for any score threshold. In fact, at equal safety (1.4% missed diagnoses among non-referred patients), the efficiency of the updated rule was lower (43.5%) compared

Table 5.5: Presence of deep vein thrombosis (DVT) in relation to the results of the original scoring rule (Table 5.1) without the D-dimer test result and to the D-dimer test result itself

	DVT absent, n	**DVT present, n**	**Total, n (% DVT)**
Clinical Characteristics only ≤ 3	767	82	849 (9.7%)
D-dimer normal	488	7	495 (1.4%)
D-dimer abnormal	279	75	354 (21.2%)
Clinical Characteristics only ≥ 4	99	54	153 (35.3%)
D-dimer normal	42	13	55 (23.6%)
D-dimer abnormal	55	40	95 (42.1%)
D-dimer unknown*	2	1	3 (33.3%)

* The contents of this table are based on the un-imputed data. All data was available apart from 3 D-dimer test results in patients with scores ≥ 4 based on only the clinical characteristics of the rule. The reason for not performing the D-dimer test was because these patients were referred anyhow as the score was already larger than 3 without the test.
n = number of patients

to the original rule (49.4%). The diagnostic contribution of the two additional predictors was apparently not around the score thresholds 3 to 5.

Our data further indicate that physicians may refrain from performing D-dimer testing in patients with a score of 4 points or more based on the clinical characteristics of the rule only (i.e. in 15% of all patients suspected of DVT). In these patients, the risk of having DVT is still high (23.6%) when the D-dimer test result is normal; they should be referred for ultrasonography regardless of their D-dimer level. To reduce both the number of ultrasounds and D-dimer tests in suspected DVT in primary care, we therefore propose the strategy as shown in Figure 5.1.

A few methodological issues should be discussed. First, different reference standards were used, depending on a patient's score. According to the protocol, patients with a score ≤ 3 were not referred for ultrasonography, in contrast to patients with scores above this threshold. However, all patients were followed for three months to document all possible cases of venous thromboembolism. Such a three month follow up period is commonly used in studies on the diagnosis of DVT.[19,20] It is assumed that if a thrombus were present at first presentation, this would have been clinically identified within this follow-up period. Therefore, bias due to differential verification (use of different reference standards) was unlinkely. Second, by using our large database of patients suspected of DVT, we tested whether the effect of the 8 predictors included in the rule should be adjusted and we explored the added predictive value of a large number (i.e. 24) other possible predictors. Since we used an alpha of < 0.05 to indicate statistical significance, which is associated with a 1 in 20 risk of finding spuriously significant results, 1 or 2 spuriously significant results can be expected. The possibility that the two significant results we found were false-positives is of limited importance only, because adding these two potential predictors to the diagnostic rule did not improve the rule's safety or efficiency anyhow.

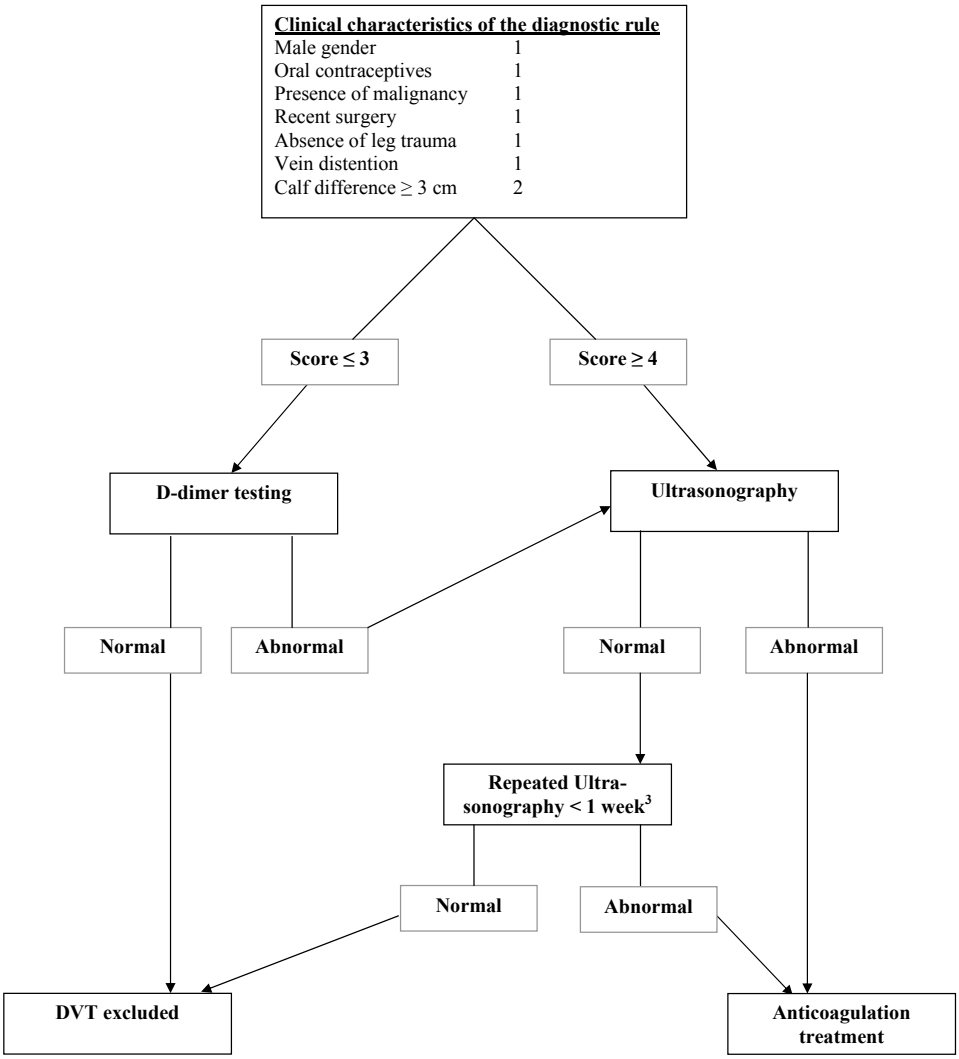

Figure 5.1: Proposed diagnostic strategy for safely excluding deep vein thrombosis in primary care.

In conclusion, the recently developed diagnostic rule for excluding DVT in primary care has optimal accuracy and does not require any adjustments. In addition to previously performed validation studies, the result of this study may provide the final evidence that the diagnostic rule can be used to safely reduce the number of ultrasound and D-dimer measurements in primary care patients clinically suspected of DVT.

Acknowledgement

We gratefully acknowledge the support by The Netherlands Organization for Scientific Research (ZonMw 016.046.360; ZonMw 945-04-009).

References

[1] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[2] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[3] Oudega R., Hoes A. W., Toll D. B., and Moons K. G. The value of clinical findings and d-dimer tests in diagnosing deep vein thrombosis in primary care, 2006, *Semin Thromb Hemost*, 32(7):673–7.

[4] AMUSE-Investigators. Safely ruling out deep venous thrombosis in primary care, 2007.

[5] Janssen K. J. M., Moons K. G. M., Kalkman C. J., Grobbee D. E., and vergouwe Y. Updating methods improved the predictive performance of clinical prediction models in new patients, 2008, *J Clin Epidemiol*, 61(1):76 – 86.

[6] Steyerberg E. W., Borsboom G. J., Houwelingen H. C., Eijkemans M. J., and Habbema J. D. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage, 2004, *Stat Med*, 23(16):2567–86.

[7] Hosmer D. W. and S L. *Applied logistic regression*. John Wiley and Sons, Inc., New York, 1989.

[8] Ivanov J., Tu J. V., and Naylor C. D. Ready-made, recalibrated, or remodeled? issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery, 1999, *Circulation*, 99 (16):2098–104.

[9] Cini M., Legnani C., Cavallaroni K., Bettini F., and Palareti G. A new rapid bedside assay for d-dimer measurement (simplify d-dimer) in the diagnostic work-up for deep vein thrombosis, 2003, *J Thromb Haemost*, 1(12):2681–3.

[10] Neale D., Tovey C., Vali A., Davies S., Myers K., Obiako M., Ramkumar V., and Hafiz A. Evaluation of the simplify d-dimer assay as a screening test for the diagnosis of deep vein thrombosis in an emergency department, 2004, *Emerg Med J*, 21(6):663–6.

[11] van der Velde E. F., Wichers I. M., Toll D. B., Weert H. C., and Buller H. R. Feasibility and accuracy of a rapid 'point-of-care' d-dimer test performed with a capillary blood sample, 2007, *J Thromb Haemost*, 5(6):1327–30.

[12] Lensing A. W., Prandoni P., Prins M. H., and Buller H. R. Deep-vein thrombosis, 1999, *Lancet*, 353 (9151):479–85.

[13] Harrell J. *Regression Modelling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York, 2001.

[14] Van Houwelingen J. C. and Le Cessie S. Predictive value of statistical models, 1990, *Stat Med*, 9(11): 1303–25.

[15] Donders A. R., Heijden G. J., Stijnen T., and Moons K. G. Review: a gentle introduction to imputation of missing values, 2006, *J Clin Epidemiol*, 59(10):1087–91.

[16] Greenland S. and Finkle W. D. A critical look at methods for handling missing covariates in epidemiologic regression analyses, 1995, *Am J Epidemiol*, 142(12):1255–64.

[17] Little R. Regression with missing x's: a review, 1992, *J Am Stat Assoc*, 87:1255–1264.

[18] van der Heijden G. J., Donders A. R., Stijnen T., and Moons K. G. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example, 2006, *J Clin Epidemiol*, 59(10):1102–9.

[19] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[20] Rathbun S. W., Whitsett T. L., and Raskob G. E. Negative d-dimer result to exclude recurrent deep venous thrombosis: a management trial, 2004, *Ann Intern Med*, 141(11):839–45.

Chapter 6

Cost-effectiveness of ruling out Deep Venous Thrombosis in primary care versus care as usual

The contents of this chapter are based on
ten Cate-Hoek AJ, Toll DB, van der Velde EF, Prins MH, Moons KGM, Oudega R, van Weert HCPM, Stoffers HEJH, Hoes AW, Buller HR, Joore MA
Cost-effectiveness of ruling out deep venous thrombosis in primary care versus care as usual
Submitted

Introduction

As deep venous thrombosis (DVT) is a potentially life threatening disorder, current primary care practice is to refer all patients suspected of DVT for objective testing to specialized diagnostic services. These services are readily available, use non-invasive tests (such as ultrasonography and D-dimer testing) and provide the referring (primary care) physician with the assurance that a case of DVT is not missed.[1,2] However, numerous studies have revealed that 80 to 90% of these referred patients do not have deep venous thrombosis.[2,3] Therefore, it would be ideal to exclude DVT at initial presentation in a large proportion of these patients and thereby avoid referral, and hence save costs.

Recently, a simple diagnostic rule was developed for general practitioners to safely exclude clinically suspected DVT (further referred to as 'AMUSE strategy').[4,5] The rule combines 7 clinical characteristics plus the result of a D-dimer test (Table 6.1), and was developed to discriminate patients unlikely to have DVT (who would not need additional ultrasonography) from patients with an increased risk of DVT requiring further work-up. With the recent introduction of so-called 'point of care' (POC) D-dimer tests, this diagnostic rule can be completed entirely in primary care. In a large (n = 1002) prospective management study, general practitioners actually used the diagnostic rule to decide which patients suspected of DVT needed referral. The number of patient referrals for ultrasound measurements was reduced by almost 50%, at the cost of an acceptably low risk (1.4%) of subsequent venous thromboembolic events in the non-referred patients.[6] Moreover, there was no need for recalbiration of the rule.[7]

The present study uses the data of the management study to estimate the cost-effectiveness of the AMUSE strategy for suspected DVT in primary care as compared to usual care strategies, using Markow modeling.

Methods

Model description

A Markov model was constructed with mutually exclusive health states (see below). The model simulated the course of events in a hypothetical cohort of 1002 persons who initially present to their primary care physician with signs and symptoms suggestive of DVT of the leg (equal to the cohort included in the prospective study). First order Monte Carlo simulation was not feasible because patient characteristic dependent probabilities were not available for the alternative diagnostic strategies. In the

Table 6.1: Diagnostic rule for diagnosing deep venous thrombosis in primary care[4]

Variable	Points
Male Gender	1
Use of hormonal contraceptives	1
Active malignancy in past 6 months	1
Recent surgery in previous month	1
Absence of leg trauma	1
Distension of collateral veins	1
Difference in calf circumference $\geq$ 3 cm	2
D-dimer assay abnormal	6

If score $\leq$ 3, no referral for ultrasound; If score $\geq$ 4, referral for ultrasound.

model, as time progresses, persons could move between the different health states according to a set of transition probabilities. The cycle length of the model was set to six months, with a 5 year time horizon. The model was constructed to compare the expected 5-year costs and health effects of different diagnostic strategies for suspected DVT. The following diagnostic strategies were compared:

1. AMUSE strategy in primary care, followed by ultrasound in hospital for patients with a score of 4 and higher (referred to as AMUSE).
2. Referral to the hospital with ultrasound for all patients (referred to as Hospital).
3. Referral to the hospital with secondary care DVT ("Wells") rule[8] in hospital followed by ultrasound for patients with a Wells score of 2 and higher (referred to as Hospital rule).

Model construction

Health states

Health states in the model were: no history of DVT, post venous thromboembolism (Post VTE), post thrombotic syndrome (PTS), and central nervous system bleeding (CNS bleed). The probability of recurrent VTE is higher for persons that have experienced a prior event of VTE compared to the population risk, therefore the health state Post VTE was included in the model. The health states PTS and CNS bleed were included because these conditions cause disutility and costs. The following events were modelled: DVT, pulmonary embolism (PE), major (gastrointestinal) bleed and CNS bleed. The final absorbing state was Death. Figure 6.1 is a graphical presentation of the model structure.

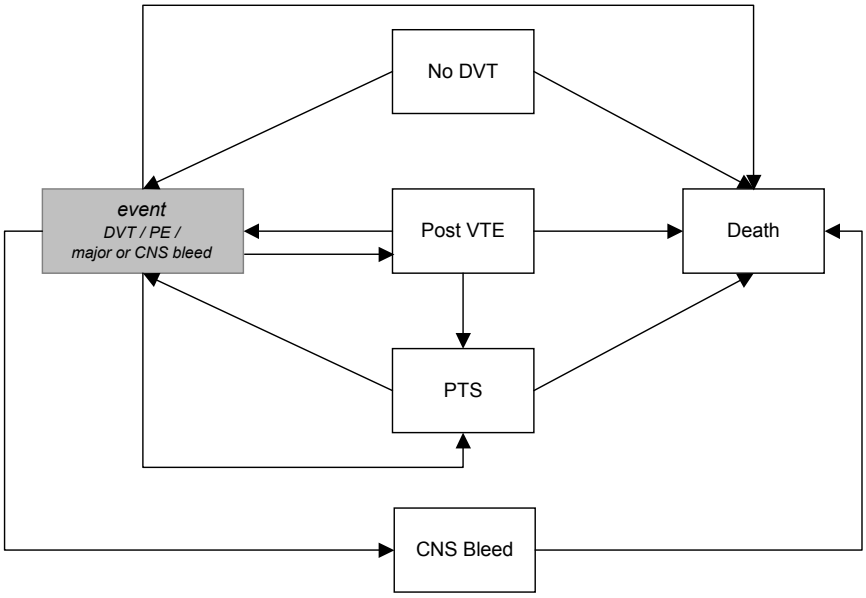

Figure 6.1: Model Structure

Model assumptions

It was assumed that every patient with DVT or recurrent episodes of DVT and PE was treated for six months with anticoagulant medication. The costs and consequences of the events DVT, PE, major bleed, and CNS bleed are calculated for six months, after this period persons move to one of the health states: post VTE, PTS, CNS bleed, or Death.

Probabilities

The probabilities associated with the diagnostic strategies and the consequences of treatment for VTE are given in Table 6.2. Data were derived from the management study (AMUSE), as well as from the literature. Background mortality was based on age-specific death rates from the Central Bureau of Statistics.

Health effects

The health state utilities assigned to the different health states and events are presented in Table 6.3. The disutility from having PTS and from experiencing an event

Table 6.2: Transition probabilities

Parameter	Mean	SE	Distribution	Source
Prevalence VTE first cycle	0,14	0,01	beta	AMUSE[6]
Ratio complaint to event	0,14	0,01	beta	AMUSE[6]
Sensitivity AMUSE strategy	0,9265	0,0223	beta	126/136; AMUSE[6]
Positive AMUSE rule	0,5010	0,0158	beta	AMUSE[6]
Sensitivity hospital strategy	0,9770	0,0160	beta	85/87; Wells, 2003[8]
Sensitivity hospital rule strategy	0,9692	0,0061	beta	787/812; Meta-analysis[a]
Positive hospital rule	0,8182	0,0164	beta	Ten Cate-Hoek, 2005[3]
PTS	0,03	0,01	beta	Prandoni, 1996[9]
Major bleed in treated patients	0,0210	0,0014	dirichlet	Linkins, 2003[10]
Fatal bleed in treated patients	0,0034	0,0006	dirichlet	Linkins, 2003[10]
CNS bleed in treated patients	0,0012	0,0003	dirichlet	Linkins, 2003[10]
Fatality of PE in treated patients	0,43	0,07	beta	Douketis, 1998[11]
Population risk event VTE	0,001	0,0001	beta	White, 2003[12]
Recurrent VTE	0,03	0,01	beta	Prandoni, 1996[9]
PE given VTE	0,27	0,13	beta	AMUSE[6]

[a] The meta-analysis of Ten Cate-Hoek, 2005[3] was extended with data from the contributing studies.[8,13–17]

DVT were calculated from EQ5D data of the management study. The quality of life of persons in the health states No DVT and Post VTE was assumed to be equal to the quality of life of persons in the general population. As utility weights, age specific EQ5D norm values for the general population were used.[18] For the events PE and major bleed and the health state CNS bleed Time Trade Off values from the literature were used.[19,20]

Costs

All unit costs were based on actual costs or standard unit costs from the Dutch Cost Manual.[21] Volumes of medical consumption were based on the prospective study, the literature, as well as on expert opinion. In Table 6.4 the mean costs for the three diagnostic strategies and for the health states and events are listed.

The costs of the strategies included the costs of medical care and travel costs to the GP and/or the hospital. In the AMUSE strategy two GP consultations, the point of care d-dimer test (Simplify®), GP time to perform d-dimer testing, and, in case of referral based on a positive rule, ER visit, ultrasound, and in-hospital lab procedures were included. In the hospital strategy only one GP consultation was included, while all patients received an ER visit and ultrasound. In the hospital rule strategy the ultrasound was limited to patients with a positive rule including a hospital d-dimer

Table 6.3: Utilities and disutilities for health states and events

Utilities	Mean	SE	Distribution	Source
Health states				
No DVT & Post VTE			beta	Age specific norm values, Kind, 1999[18]
18-19	0.94	0.02		
20-24	0.94	0.02		
25-29	0.93	0.03		
30-34	0.93	0.03		
35-39	0.91	0.04		
40-44	0.91	0.04		
45-49	0.85	0.04		
50-54	0.85	0.04		
55-59	0.80	0.04		
60-64	0.80	0.04		
65-69	0.78	0.04		
70-74	0.78	0.04		
75-79	0.73	0.04		
Disutility PTS	0.02	0.03	beta	AMUSE[6]
CNS Bleed	0.33	0.01	beta	Van Dongen, 2004[19]
Events				
Deep Venous Thrombosis	0.67	0.03	beta	AMUSE[6]
Pulmonary Embolism	0.62	0.01	beta	Van Dongen, 2004[19]
Major bleed	equal to pulmonary embolism			Based on Locadia, 2004[20]

test. It was assumed that persons in the health states No DVT and Post VTE did not experience any costs related to DVT. Persons in the health state PTS and CNS bleed did experience costs. To calculate the costs of diagnosing DVT after the initial presentation of complaints (in the following Markov cycles), the incidence of VTE as found in the literature was multiplied by the ratio complaint to documented VTE as observed in the prospective study. Details of the costs calculations are listed in the Appendix.

Analysis

We compared the cost-effectiveness of three diagnostic strategies: the AMUSE strategy, the hospital strategy and the hospital rule strategy. Incremental cost-effectiveness ratios (iCERs) were calculated, dividing the incremental costs by the incremental QALYs. ICERs were calculated by comparing each strategy with the next most effective strategy. Whether a strategy is deemed efficient depends on how much society is willing to pay for a gain in effect, which is referred to as the ceiling ratio. In the Netherlands an informal ceiling ratio of €80,000 per QALY exists.[23] This is however a maximum ceiling ratio which applies when there is a high burden of disease. Al-

Table 6.4: Summary of cost parameters per cycle of 6 months or per event

Parameter	Mean value (Euro)	Sources	Uncertainty
Diagnostic strategies			
AMUSE strategy	168	Various, see appendix	Fixed
Hospital strategy	251	Various, see appendix	Fixed
Hospital rule strategy	227	Various, see appendix	Fixed
Travel for diagnosis			
Travel to GP	3	Various, see appendix	See appendix
Travel to hospital	7	Various, see appendix	See appendix
Health states			
PTS	3,247	Various, see appendix	minimum €140, maximum €10,580*
CNS bleed	28,419	Costs of nursing home admission	Fixed
Events			
Incident PTS	3,367	Various, see appendix	minimum €273, maximum €10,6701
DVT	1,322	Various, see appendix	See appendix
PE	4,210	Various, see appendix	See appendix
Major bleed	4,211	Various, see appendix	minimum €1688, maximum €11,4971
CNS bleed	11,281	Bergman et al, 1995 [22]	Fixed

* A beta pert distribution was used in the probabilistic sensitivity analysis

though this, may not directly apply to DVT, the complications of (missed) DVT, such as PE and PTS, and the side-effects of the treatment of DVT, like CNS bleed, can be considered as serious conditions. The National Institute for Health and Clinical Excellence in the United Kingdom uses a ceiling ratio between £20,000 and £30,000 per QALY,[24] which is roughly €40,000.

Uncertainty surrounding the iCERs was handled probabilistically. This means that we assigned distributions to the model parameters, to reflect the second-order uncertainty in the estimation of that parameter.[25] Measures of variance were retrieved from the prospective study, the patient cohort or published literature and, if no other source was available, from expert opinion. See Table 6.2 for the assigned distributions. Parameter values were drawn at random from the assigned distributions, using Monte Carlo simulation with 1000 iterations. To illustrate the results of the simulation, cost-effectiveness acceptability curves (CEACs) were calculated.[26,27] For different ceiling ratios, the net monetary benefit was calculated for each strategy by subtracting the costs from the effects, multiplied by the ceiling ratio. CEACs show the probability that a pathway has the highest net monetary benefit, and thus is deemed cost-effective, given different ceiling ratios.

Costs were calculated to their 2004 value using price index figures form the Cen-

tral Bureau of Statistics. Future costs and effects were discounted to their present value by a rate of 4% and 1.5% respectively, according to Dutch guidelines.[28] The base case analysis was based on the observed age (58 years) and sex distribution (375 males; 37%) as well as the proportion of patients with PTS (62; 6%) and previous VTE (170; 17%) in the prospective study.

Sensitivity analyses were performed to test the consistency of the results. Included in the one-way sensitivity analyses were the discount rate and age, as well as the following (partially) fixed cost parameters: the health states PTS and CNS bleed, and the events DVT, PE, major bleed and CNS bleed. Furthermore, we calculated the costs and the sensitivity of the AMUSE diagnostic strategy for iCERs of €40,000 per QALY.

Results

Base case analysis

The AMUSE strategy had both slightly lower costs and less QALYs than both care as usual strategies. The hospital strategy was the most effective, and had the highest costs. The iCER of the hospital strategy versus the hospital rule strategy amounts to €91,057. This indicates that, even based on a maximum threshold of €80,000, the hospital rule strategy is to be preferred. Therefore, we compared the AMUSE strategy to the hospital rule strategy. This resulted in a cost saving of €138, and a QALY loss of 0,002. The iCER is €55,753. If usual care consists of an equal mix of the hospital and hospital rule strategy, the iCER of the AMUSE strategy is €58,662. See Table 6.5.

The probabilistic sensitivity analysis showed that the simulation results of the cost

Table 6.5: Results of the cost-effectiveness analysis

Diagnostic strategy	**Life Years**	**QALYs**	**Costs (Euros)**
AMUSE	4.8723	3.8532	3,589
Hospital rule	4.8774	3.8557	3,727
Hospital	4.8782	3.8562	3,768
Combination of hospital (50%) and hospital rule (50%)	4.8778	3.8559	3,747

and QALY outcomes of the diagnostic strategies were comparable (data not shown). The incremental costs and QALYs from the Monte Carlo simulation comparing the AMUSE strategy and the hospital rule strategy are all in the south of the cost-effectiveness plane, and mostly in the southwest quadrant (Figure 6.2). This indicates cost savings

for (in majority) a QALY loss. The cost-effectiveness acceptability curves (Figure 6.3) show that the hospital strategy has the lowest probability of being cost-effective for a threshold of the iCER up to €68,000. The AMUSE strategy has the highest probability of being cost-effective as long as society demands less than €100,000 compensation for a QALY loss. For a threshold of €40,000, the probability that the AMUSE strategy is cost-effective is 72%, while for a threshold of €80,000 it is still 44%.

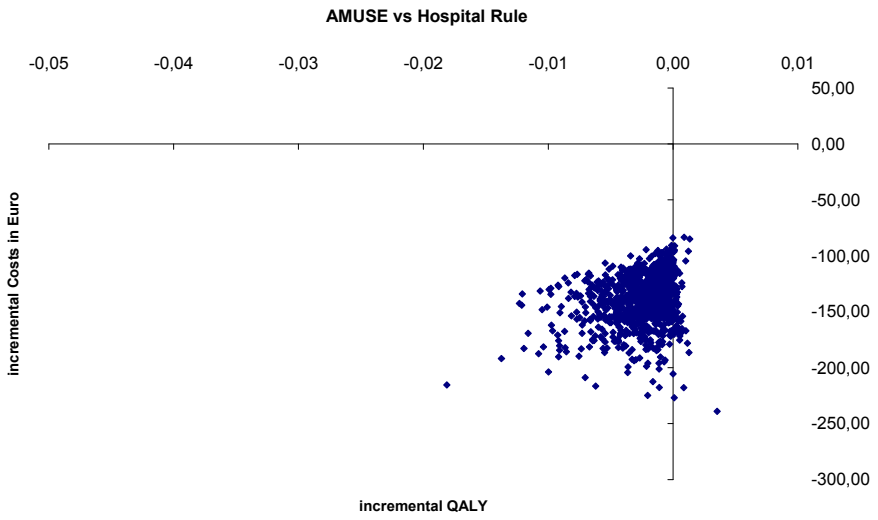

Figure 6.2: Cost-effectiveness plane

Additional sensitivity analyses

The discounting rate and age did not influence the result. Across a range of 30 to 80 years the iCER of the AMUSE strategy versus the hospital rule strategy varied with only €1.300 (data not shown). If the sensitivity of the AMUSE strategy was decreased from 0.9265 to 0.9032 the iCER amounted to €40,000. An increase of the costs of the AMUSE diagnostic strategy of €27 to €195 per patient, resulted in an iCER of €40,000. A wide range of variation in the costs for health states and events did not change our results substantially. The results of the sensitivity analyses are presented in Table 6.6.

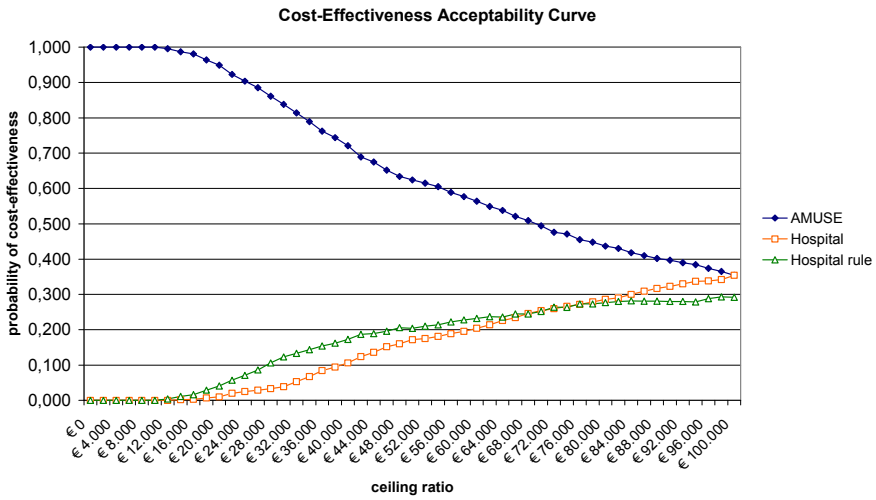

Figure 6.3: Cost-effectiveness acceptability curves

Discussion

The model based cost-effectiveness analysis showed that the AMUSE strategy,[4,5] which was deemed to be safe clinically,[6] was associated with a minimal (less than 0.01) loss of QALYs. Also, the cost saving associated with the strategy was relatively small (€138). This resulted in a marginally acceptable iCER of the AMUSE strategy versus the next best alternative, the hospital rule strategy (€55,753/QALY). However, the cost-effectiveness acceptability curves show that the AMUSE strategy has a higher probability of being cost-effective than the alternative strategies for thresholds up to €100,000. For a threshold of €40,000, the probability that the AMUSE strategy is cost-effective is 72%. Hence, it can be concluded that the AMUSE strategy for the diagnosis of suspected deep vein thrombosis with a point of care d-dimer test combined with a clinical decision rule in primary care is cost-effective.

Some details of our model study require attention. Some costs of health states and events were estimates based on expert opinion or are partially based on assumptions. However, the sensitivity analyses show that for a wide range of variation in these costs our results do not change substantially. The cost-effectiveness outcome is strongly influenced by the sensitivity of the diagnostic strategies. While the sensitivity of the hospital rule strategy was based on a meta-analysis including over 6000 patients, the estimates for the hospital and AMUSE strategies were based on studies including 1100 and 1000 patients respectively. This resulted in larger uncertainty for the sensitivity of

Table 6.6: Results of the sensitivity analysis

Parameter in the sensitivity analysis	Diagnostic strategy	QALYs	Costs Euro	iCER Euro/QALY
Base case	**AMUSE**	**3.8532**	**3,589**	**55,753**
	Hospital rule	**3.8557**	**3,727**	
Undiscounted	AMUSE	3.8820	3,660	56,436
	Hospital rule	3.8845	3,801	
Sensitivity AMUSE strategy	AMUSE	3.8519	3,574	40,000
0.9032 instead of 0.9265	Hospital rule	3.8557	3,727	
Costs AMUSE strategy	AMUSE	As in	3,628	40,000
€195 instead of €168	Hospital rule	base case	3,727	
Costs of PTS high value	AMUSE		9,919	68,767
€10,580 instead of €3,247	Hospital rule	As in	10,089	
Costs of PTS low value	AMUSE	base case	907	50,240
€140 instead of €3,247	Hospital rule		1,031	
Costs of CNS bleed high value	AMUSE		3,637	56,706
€56,838 instead of €28,419	Hospital rule	As in	3,777	
Costs of CNS bleed low value	AMUSE	base case	3,565	55,277
€14,210 instead of €28,419	Hospital rule		3,701	
Costs of PTS incident high value	AMUSE		4,025	58,972
€10,670 instead of €3,367	Hospital rule	As in	4,171	
Costs of PTS incident low value	AMUSE	base case	3,404	54,466
€274 instead of €3,367	Hospital rule		3,539	
Costs of event DVT high value	AMUSE		3,819	60,499
€2,644 instead of €1,322	Hospital rule	As in	3,969	
Costs of event DVT low value	AMUSE	base case	3,474	53,379
€661 instead of €1,322	Hospital rule		3,606	
Costs of event PE high value	AMUSE		3,607	51,427
€8,420 instead of €4,210	Hospital rule	As in	3,735	
Costs of event PE low value	AMUSE	base case	3,580	57,916
€2,104 instead of €4,210	Hospital rule		3,723	
Costs of event major bleed high value	AMUSE		3,620	56,386
€11,497 instead of €4,211	Hospital rule	As in	3,759	
Costs of event major bleed low value	AMUSE	base case	3,578	55,534
€2,104 instead of €1,688	Hospital rule		3,716	
Costs of event CNS bleed high value	AMUSE		3,592	55,810
€22,562 instead of €11,281	Hospital rule	As in	3,730	
Costs of event CNS bleed low value	AMUSE	base case	3,588	55,725
€5,641 instead of €11,281	Hospital rule		3,725	

these strategies. In current practice it is likely that hospitals use a mix of the hospital and hospital rule strategy. Therefore, our results can be considered as conservative. Finally, it could be argued that results would be more favourable if a more sensitive d-dimer test combined with the decision rule would have been used in the AMUSE strategy. However, in that case also the specificity is likely to be lower resulting in more patients referred to secondary care, and hence less cost savings.

For clinical practice, implementation of the AMUSE strategy results in the exclusion of DVT in approximately 50% of patients in primary care.[6] This may have the added benefit of convenience for the patients, since referral for ultrasound is not necessary, and may enable the general practitioner to direct attention at finding alternative diagnoses without delay. Although the general practitioner, or the practice assistant, spends extra time to perform the d-dimer test and apply the rule (for which extra costs were included in our model), the prospective study proved that implementation in general practice was highly feasible. In order to facilitate successful implementation a reimbursement for the extra time of the general practitioner would be appropriate.

In summary, the AMUSE strategy to exclude DVT in primary care is not only safe, but also cost-effective as compared to hospital based strategies, with or without the use of a diagnostic rule, to diagnose DVT.

Acknowledgement

We gratefully acknowledge the support by The Netherlands Organization for Scientific Research (ZonMw 945-04-009).

References

[1] Kraaijenhagen R. A., Lensing A. W., Lijmer J. G., Prandoni P., Prins M. H., Ginsberg J. S., and Buller H. R. Diagnostic strategies for the management of patients with clinically suspected deep-vein thrombosis, 1997, *Curr Opin Pulm Med*, 3(4):268–74.

[2] Lensing A. W., Prandoni P., Prins M. H., and Buller H. R. Deep-vein thrombosis, 1999, *Lancet*, 353 (9151):479–85.

[3] Ten Cate-Hoek A. J. and Prins M. H. Management studies using a combination of d-dimer test result and clinical probability to rule out venous thromboembolism: a systematic review, 2005, *J Thromb Haemost*, 3(11):2465–70.

[4] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[5] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[6] AMUSE-Investigators. Safely ruling out deep venous thrombosis in primary care, 2007.

[7] Toll D. B. and colleagues. Optimization of the diagnostic strategy for suspected deep vein thrombosis in primary care (chapter 5 of this thesis), 2008.

[8] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[9] Prandoni P., Lensing A. W., Cogo A., Cuppini S., Villalta S., Carta M., Cattelan A. M., Polistena P., Bernardi E., and Prins M. H. The long-term clinical course of acute deep venous thrombosis, 1996, *Ann Intern Med*, 125(1):1–7.

[10] Linkins L. A., Choi P. T., and Douketis J. D. Clinical impact of bleeding in patients taking oral anticoagulant therapy for venous thromboembolism: a meta-analysis, 2003, *Ann Intern Med*, 139 (11):893–900.

[11] Douketis J. D., Kearon C., Bates S., Duku E. K., and Ginsberg J. S. Risk of fatal pulmonary embolism in patients with treated venous thromboembolism, 1998, *Jama*, 279(6):458–62.

[12] White R. H. The epidemiology of venous thromboembolism, 2003, *Circulation*, 107(23 Suppl 1): I4–8.

[13] Bates S. M., Kearon C., Crowther M., Linkins L., O'Donnell M., Douketis J., Lee A. Y., Weitz J. I., Johnston M., and Ginsberg J. S. A diagnostic strategy involving a quantitative latex d-dimer assay reliably excludes deep venous thrombosis, 2003, *Ann Intern Med*, 138(10):787–94.

[14] Anderson D. R., Kovacs M. J., Kovacs G., Stiell I., Mitchell M., Khoury V., Dryer J., Ward J., and Wells P. S. Combined use of clinical assessment and d-dimer to improve the management of patients presenting to the emergency department with suspected deep vein thrombosis (the edited study), 2003, *J Thromb Haemost*, 1(4):645–51.

[15] Schutgens R. E., Ackermark P., Haas F. J., Nieuwenhuis H. K., Peltenburg H. G., Pijlman A. H., Pruijm M., Oltmans R., Kelder J. C., and Biesma D. H. Combination of a normal d-dimer concentration and a non-high pretest clinical probability score is a safe strategy to exclude deep venous thrombosis, 2003, *Circulation*, 107(4):593–7.

[16] Kearon C., Ginsberg J. S., Douketis J., Crowther M., Brill-Edwards P., Weitz J. I., and Hirsh J. Management of suspected deep venous thrombosis in outpatients by using clinical assessment and d-dimer testing, 2001, *Ann Intern Med*, 135(2):108–11.

[17] Janes S. and Ashford N. Use of a simplified clinical scoring system and d-dimer testing can reduce the requirement for radiology in the exclusion of deep vein thrombosis by over 20

[18] Kind P., Hardman G., and Macran S. *UK population norms for EQ-5D: discussion paper 172*. The University of York, Centre for Health Economics, York, 1999.

[19] Dongen C. J. *An evidence based approach to optimizing anticoagulant strategies*. PhD thesis, Universiteit van Amsterdam, 2004.

[20] Locadia M., Stalmeier P. F., Oort F. J., Prins M. H., Sprangers M. A., and Bossuyt P. M. A comparison of 3 valuation methods for temporary health states in patients treated with oral anticoagulants, 2004, *Med Decis Making*, 24(6):625–33.

[21] Oostenbrink J. B., Koopmanschap M. A., and Rutten F. F. G. *Dutch cost manual (in Dutch)*. College voor Zorgverzekeringen, Amstelveen, 2004.

[22] Bergman L., Meulen J. H., Limburg M., and Habbema J. D. Costs of medical care after first-ever stroke in the netherlands, 1995, *Stroke*, 26(10):1830–6.

[23] *Sensible and sustainable care (in Dutch)*. Council for Public Health and Health Care, Zoetermeer, 2006.

[24] Buxton M. J. Economic evaluation and decision making in the uk, 2006, *Pharmacoeconomics*, 24(11): 1133–42.

[25] Weinstein M. C. Recent developments in decision-analytic modelling for economic evaluation, 2006, *Pharmacoeconomics*, 24(11):1043–53.

[26] Fenwick E., Claxton K., and Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves, 2001, *Health Econ*, 10(8):779–87.

[27] Hout B. A., Al M. J., Gordon G. S., and Rutten F. F. Costs, effects and c/e-ratios alongside a clinical trial, 1994, *Health Econ*, 3(5):309–19.

[28] *Health Care Insurance Board*. 2006.

[29] Caprini J. A., Botteman M. F., Stephens J. M., Nadipelli V., Ewing M. M., Brandt S., Pashos C. L., and Cohen A. T. Economic burden of long-term complications of deep vein thrombosis after total hip replacement surgery in the united states, 2003, *Value Health*, 6(1):59–74.

Appendix: Details of cost calculations (euro's, 2004)

Table 6.7: Costs of the diagnostic strategies

Parameter	Mean value	Source
AMUSE strategy	167,82	
At GP		
GP consultation	20,60	Dutch Cost Manual [21]
# of GP consultations	2,00	AMUSE [6]
Costs of d-dimer	9,25	Company
Costs of performing d-dimer	10,30	Dutch Cost Manual [21]
In hospital		
Costs of blood draw	10,64	Dutch Cost Manual [21]
Costs of a lab procedure	1,78	Dutch Cost Manual [21]
# lab procedures	5,00	Haematology (3), clinical chemistry (2); expert opinion
Costs of 1 ER visit	141,78	Dutch Cost Manual [21]
Costs per ultrasound	40,69	Dutch Cost Manual [21]
# of ultrasounds rule	1,29	Wells, 2003 [12]
Hospital strategy	251,08	
# of GP consultations	1,00	assumption
# lab procedures	5,00	Haematology (3), clinical chemistry (2); expert opinion
Costs of 1 ER visit	141,78	Dutch Cost Manual [21]
# of ultrasounds	1,70	Kraaijenhagen, 1997 [1]
Hospital rule strategy	226,54	
# of GP consultations	1,00	assumption
Blood draw	10,64	Dutch Cost Manual [21]
# lab procedures & d-dimer	6,00	Haematology (3), clinical chemistry (2), d-dimer; expert opinion
Costs of 1 ER visit	141,78	Dutch Cost Manual [21]
# of ultrasounds	1,29	Wells, 2003 [8]
Travel for diagnosis		
Travel to GP	2,85	
Travel to hospital	7,38	
GP km	1,80	Dutch Cost Manual [21]
hospital km	7,00	Dutch Cost Manual [21]
Costs of km per car	0,16	Dutch Cost Manual [21]
Costs of car parking	2,55	Dutch Cost Manual [21]
GP, proportion by car	0,50	AMUSE [6], se 0,02, beta distribution
Hospital, proportion by car	1,00	AMUSE [6]

Table 6.8: Costs associated with events

Parameter	Mean value	se	Distribution	Source
Event deep venous thrombosis	1322,45			
GP consultation	29,60			Dutch Cost Manual [21]
# of GP consultations	0,83	0,30	gamma	AMUSE [6]
Home care compression therapy	480,62			AMUSE, Dutch Cost Manual [6,21]
LMWH 7 days	66,68			Pharmacotherapeutic Compass
Coumarins 6 months	85,61			Pharmacotherapeutic Compass
specialist visit	57,12			Dutch Cost Manual [21]
# control visits specialist	2,79	0,84	gamma	AMUSE [6]
INR control visit	8,46			Thrombosis Service
# INR control visits	16,38	1,28	gamma	AMUSE [6]
compression stockings	60,38			Health care insurance company
Hospital day	485,52			Dutch cost manual [21]
# hospital days	0,63	0,11	gamma	AMUSE [6]
Event pulmonary embolism	4209,77			
GP consultation	29,60			Dutch Cost Manual [21]
# of GP consultations	1,42	1,07	gamma	AMUSE [6]
ER visit	141,78			Dutch Cost Manual [21]
CT thorax	132,35			Dutch Cost Manual [21]
ECG	25,41			Dutch Cost Manual [21]
Blood draw	10,64			Dutch Cost Manual [21]
Lab procedures	1,78			Dutch Cost Manual [21]
# lab procedures	5,00			Haematology (3), clinical chemistry (2); expert opinion
Hospital day	485,52			Dutch Cost Manual [21]
# hospital days	7,00			Expert opinion
LMWH 7 days	66,68			Pharmacotherapeutic Compass
Coumarins 6 months	85,61			Pharmacotherapeutic Compass
# control visits specialist	2,79	0,84	gamma	AMUSE [6]
# INR control visits	16,38	1,28	gamma	AMUSE [6]

Table 6.9: Incidence and health state costs associated with Post Thrombotic Syndrome

Parameters	Proportion of patients [29]	Number [29]	Unit costs Euro	Source	Costs Euro
Mild to moderate first 6 months					
Specialist contact	1	2,00	57,12	Dutch Cost Manual [21]	114,24
Duplex	1	1,00	90,11	Dutch Cost Manual [21]	90,11
Stockings	1	1,00	60,38	Insurance Company	60,38
Vein ligation and stripping	0,015	1,00	609,90	Dutch Cost Manual [21]	9,15
Total					273,87
Mild to moderate after 6 months					
Specialist contact	1	2,00	57,12	Dutch Cost Manual [21]	57,12
Duplex	0,25	1,00	90,11	Dutch Cost Manual [21]	22,53
Stockings	1	1,00	60,38	Insurance Company	60,38
Total					140,02
Severe first 6 months					
Specialist contact	1	2,00	57,12	Dutch Cost Manual [21]	114,24
Duplex	1	1,00	609,90	Dutch Cost Manual [21]	90,11
Pneumatic compressor	0,07	1,00	61,01	Dutch Cost Manual [21]	4,27
Stockings	1	1,00	60,38	Insurance Company	60,38
Steroid creams (locoid crème)	0,1		9,92		0,99
Antibiotic (ciprofloxacine, 10 days bid)	0,75		32,99		24,74
Hospital days for ulcers	0,15	14,5	485,52	Dutch Cost Manual [21]	1.056,01
Home care for ulcers	0,85	182,5	60,08	Dutch Cost Manual [21]	9.319,60
Total					10.670,34
Severe after 6 months					
Specialist contact	1	2,00	57,12	Dutch Cost Manual [21]	114,24
Pneumatic compressor	0,07	1,00	61,01	Dutch Cost Manual [21]	4,27
Stockings	1	1,00	60,38	Insurance Company	60,38
Steroid creams (locoid crème)	0,1		9,92	Pharmacotherapeutic Compass	0,99
Antibiotic (ciprofloxacine, 10 days bid)	0,75		32,99	Pharmacotherapeutic Compass	24,74
Hospital days for ulcers	0,15	14,5	485,52	Dutch Cost Manual [21]	1.056,01
Home care for ulcers	0,85	182,5	60,08	Dutch Cost Manual [21]	9.319,60
Total					10.580,23
Weigthed average moderate and severe first 6 months (costs of incident)					
moderate	0,70	273,87			191,71
severe	0,30	10.670,34			3.175,69
Total					3.367,41
Weigthed average moderate and severe after 6 months (costs of health state)					
moderate	0,70	140,02			98,02
severe	0,30	10.580,23			3.148,88
Total					3.246,89

Table 6.10: Costs associated with event major bleed

Major bleed tractus digestives	**Expert 1 hematologist**		**Expert 2 hematologist**		**Expert 3 vascular internist**		**Expert 4 gastroenterologist**	
	min	max	min	max	min	max	min	max
Days of hospital admission, #	3	21	2	2	3	10	4	10
€	1.456,56	10.195,92	971,04	971,04	1.456,56	4.855,20	1.942,08	4.855,20
mean all experts, €	3.337,95							
max all experts, €	10.195,92							
min all experts, €	971,04							
Diagnostic imaging								
- gastroscopy, #	1	2	1	1	1	2	2	3
€	144,50	289,01	144,50	144,50	144,50	289,01	289,01	433,51
mean all experts, €	234,82							
max all experts, €	433,51							
min all experts, €	144,50							
- CT scan, #	0	0	1	1	0	0	0	0
€	0,00	0,00	132,35	132,35	0,00	0,00	0,00	0,00
mean all experts, €	33,09							
max all experts, €	132,35							
min all experts, €	0,00							
Lab tests, #	3	20	2	2	3	5	2	8
€	37,29	188,30	28,41	28,41	37,29	55,06	28,41	81,71
mean all experts, €	60,61							
max all experts, €	188,30							
min all experts, €	28,41							
Medication								
- Vitamin K (konakion)	1		1				1	3
one dose, €	1,02		1,02				1,02	3,06
mean all experts, €	1,53							
max all experts, €	3,06							
min all experts, €	1,02							
- Proton pump inhibitor (omeprazol)	1				1		1	
6 months, €	226,20				226,20		226,20	
mean all experts, €	226,20							
max all experts, €	226,20							
min all experts, €	226,20							
Care after hospital admission								
INR control visit	x		x				x	
24 visits, €	202,98							
Mean all experts, €	202,98							
outpatient clinic	x		x		x		x	
2 visits, €	114,24							
Mean all experts, €	114,24							
Mean sumscore, €	4.211,41							
Max sumscore, €	11.496,56							
Min sumscore, €	1.688,39							

Part II

D-dimer in suspected Deep Vein Thrombosis

Chapter 7

Different cut-off values for two D-dimer assays to exclude deep venous thrombosis in primary care

The contents of this chapter are based on
Oudega R, Toll DB, Bulten RJ, Hoes AW, Moons KG
Different cut-off values for two D-dimer assays to exclude deep venous thrombosis in primary care
Thrombosis Haemostasis 2006;95:744-6

Introduction

Deep venous thrombosis (DVT) is a serious and common disease.[1] When left untreated, it is associated with an increased risk of pulmonary embolism, whereas false positive diagnosis can lead to unnecessary anticoagulant therapy with risk of bleeding.[2,3] Diagnostic assessment of patients with suspicion of DVT can be significantly improved by additional use of D-dimer testing since patient history and physical examination alone have limited value in diagnosing deep vein thrombosis.[4–6]

D-dimer assays are commonly very sensitive and therefore suited to exclude DVT. However, there is no standardisation in the results of these assays. As a consequence, manufacturers recommend that each assay must be validated in the population at hand to determine for that population the optimal cut-off value for exclusion of DVT. This limits the comparison of different D-dimer tests across populations. A few studies evaluated the optimal thresholds for different D-dimer assays, but only in small series or in secondary care (out)patients.[7–9] There is no evidence on the optimal threshold for excluding DVT in primary care patients. As a part of a large prospective study in primary care patients suspected of DVT, we determined the performance of two often used D-dimer assays for excluding DVT.[10] Both manufacturers and investigators often consider the D-dimer test normal if the detected concentration is < 500 ng/ml.[11–13] However, little evidence is available if this threshold is to be preferred to exclude DVT, and whether different D-dimer assays require different thresholds. We evaluated two widely used sensitive D-dimer assays, a rapid ELISA method (VIDAS, bioMérieux, France) and a quantitative latex assay method (Tinaquant, Roche, Germany). Both manufacturers address 500 ng/ml as cut-off value but carefully add that each laboratory should establish its own normal range based upon the population tested. We investigated for both assays if this threshold for excluding DVT is of value in clinical practice and whether other thresholds might be preferred.

Methods

We analysed the data from a prospective study among 110 primary care physicians who investigated the diagnostic value of various diagnostics including different D-dimer assays in a large population of patients suspected of DVT.[10] In brief, the study included adult (over 18 years) patients who consulted their primary care physician with signs or symptoms suspect for DVT, defined by the presence of swelling, redness and/or pain of the legs. Patients were excluded if these symptoms existed for more than 30 days and if there was a suspicion of pulmonary embolism. After the history

had been taken and a physical examination had been performed, all patients underwent D-dimer testing performed by either the VIDAS or Tinaquant assay, depending on the laboratory routine of the respective hospital. D-dimer concentration was measured immediately after venous blood was drawn into sodium citrate tubes, using the Tinaquant D-dimer assay on a Hitachi 917 machine or for the VIDAS D-dimer assay with fluorescence detection performed on the VIDAS immunoanalyzer. There were three local hospitals involved, with a total of two laboratories. In one laboratory the VIDAS assays were tested on the stand alone machine from bioMérieux with daily control tests. The other laboratory tested the Tinaquant assays for two local hospitals on a Hitachi 917 machine with the reagent of Roche. The recommended daily controls were always done. Hence, there were no inter-laboratory or -machine variances. Finally, each patient underwent real time B-mode compression ultrasonography (CUS) with a 5–7.5 MHz linear-array sonographic scanner as reference standard. The test was repeated after 7 days in patients with a normal CUS. Proximal DVT was considered present if the result of one of the two ultrasound tests was abnormal. The sonographer was blinded to the patients' history, physical examination and D-dimer results. In relation to the results of the ultrasonography, sensitivity, specificity, negative predictive value (NPV) and negative likelihood ratio (LR-) of both D-dimer assays were calculated at different thresholds.

Results & Comments

Ultrasonography revealed DVT in 164 of the 852 analysed patients, 67 in the group that underwent the VIDAS assay (n = 425, prevalence 15.8%) and 97 in the Tinaquant group (n = 427, prevalence 22.7%). The diagnosis DVT was confirmed in 161 patients during the first ultrasound and in 3 patients at the one week repeated ultrasound. At the 500 ng/ml threshold, the VIDAS assay had a sensitivity of 99% (Table 7.1) compared to a sensitivity of 91% in the Tinaquant assay (Table 7.2). At the same threshold VIDAS had a specificity of 29%, lower than Tinaquant which scored 51%. The NPV and LR- for VIDAS were subsequently 99% and 0.05 compared to 95% and 0.18 for Tinaquant. To safely rule out DVT, a high sensitivity and NPV and a low LR- are needed to limit missed DVT cases. Likelihood ratios less than 0.1 result in large and often conclusive changes from pre- to post-test probability.[14] At the 500 ng/ml threshold, the VIDAS assay seems to meet these criteria in contrast to the Tinaquant test.

Given the main purpose of D-dimer testing - i.e. to exclude DVT - we confirm the cut-off value of 500 ng/ml for the VIDAS assay for use in primary care. In contrast,

Table 7.1: Diagnostic accuracy of the VIDAS assay at different cut-off values. To improve the readability not all 95% confidence intervals are given.

D-dimer cut-off value (ng/ml)	**DVT absent***	**DVT present***	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**NPV (95% CI)**	**LR- (95% CI)**
100	5	1	99	1	83	1.00
200	23	1	99 (96 - 100)	6 (4 - 9)	96 (79 - 100)	0.23 (0.03 - 1.60)
300	47	1	99 (96 - 100)	13 (10 - 17)	99 (89 - 100)	0.11 (0.11 - 0.12)
400	69	1	99 (96 - 100)	19 (15 - 23)	99 (92 - 100)	0.08 (0.09 - 0.08)
500	103	1	99 (96 - 100)	29 (24 - 34)	99 (95 - 100)	0.05 (0.05 - 0.05)
600	123	2	97	34	98	0.09
700	141	2	97	39	99	0.08
800	156	2	97 (93 - 100)	44 (38 - 49)	99 (96 - 100)	0.07 (0.07 - 0.07)
900	173	3	96	48	98	0.09
1000	199	8	88	56	96	0.21
1100	220	8	88	62	97	0.19
1200	241	9	87	67	96	0.20
1300	254	10	85	71	96	0.21
1400	264	10	85	74	96	0.20
1500	277	11	84	77	96	0.21
all patients	**358**	**67**				

* Number of patients below cut-off points; DVT = Deep Venous Thrombosis; NPV = negative predictive value; LR- = negative likelihood ratio; 95% CI = 95% Confidence Interval

Table 7.2: Diagnostic accuracy of the Tinaquant assay at different cut-off values. To improve the readability not all 95% confidence intervals are given.

D-dimer cut-off value (ng/ml)	**DVT absent***	**DVT present***	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**NPV (95% CI)**	**LR- (95% CI)**
100	24	1	99	7	96	0.14
200	68	2	98 (95 - 100)	21 (16 - 25)	97 (95 - 100)	0.10 (0.02 - 0.40)
300	109	5	95 (90 - 99)	33 (28 - 38)	96 (90 - 99)	0.16 (0.15 - 0.17)
400	136	6	94 (89 - 99)	41 (36 - 47)	96 (94 - 98)	0.15 (0.14 - 0.16)
500	167	9	91 (85 - 97)	51 (45 - 56)	95 (91 - 98)	0.18 (0.17 - 0.20)
600	189	11	89	57	95	0.20
700	202	12	88	61	94	0.20
800	216	14	84 (76 - 91)	65 (60 - 70)	93 (89 - 96)	0.25 (0.22 - 0.30)
900	220	19	80	67	92	0.29
1000	227	22	77	69	91	0.33
1100	230	23	76	70	91	0.34
1200	235	24	75	71	91	0.35
1300	244	27	72	74	90	0.38
1400	251	28	71	76	90	0.38
1500	258	29	70	78	90	0.38
all patients	**330**	**97**				

* Number of patients below cut-off points; DVT = Deep Venous Thrombosis; NPV = negative predictive value; LR- = negative likelihood ratio; 95% CI = 95% Confidence Interval

for the Tinaquant assay our data suggest to use a lower threshold than 500 ng/ml. The 500 ng/ml threshold yields an unacceptably low sensitivity and NPV, and a too high LR-. Cut-off values in the range of 200 to 400 ng/ml threshold, depending on whether the assay is used as stand-alone test or in combination with clinical characteristics, yielded better diagnostic accuracy. As far as we know, these data are the first from investigations in primary care. For comparison purpose, in secondary care (out)patients, van der Graaf et al.[9] and Schutgens et al.[8] found for the VIDAS assay similar sensitivity (100%), but different specificity (8% and 41%, respectively) and for the Tinaquant assay they found a sensitivity of 99% and 100%, respectively, with almost the same specificity (39% and 41%, respectively) for the cut-off value of 500 ng/ml. Schutgens et al.[8] analysed lower thresholds of both assays, and found, e.g. at a threshold of 200 ng/ml, high sensitivity (100%) but very low specificity (0–4%) for both assays. De Monye et al.[7] studied the accuracy of the VIDAS and the Tinaquant assay in patients with suspicion of pulmonary embolism (PE) in secondary care patients. At a cut-off value of 500 ng/ml, both assays showed in patients with PE lower sensitivity (88% and 82%, respectively) compared to our patients with DVT (99% and 91%). De Monye et al.[7] suggested also to lower the thresholds of both assays to improve their sensitivity and specificity. At a threshold of 200 ng/ml, they found in patients with PE an improved sensitivity of 96% for both the VIDAS and the Tinaquant assay with a specificity of 24% and 21%, respectively.

In conclusion, our study confirms that to exclude DVT in primary care, a threshold of 500 ng/ml should be used for the VIDAS assay, while for the Tinaquant assay a lower cut-off point is probably of more clinical value. Further research is needed to confirm whether for the Tinaquant assay indeed lower thresholds are preferred for excluding DVT in primary care. Care should be taken when extrapolating our thresholds and corresponding diagnostic accuracy measures to secondary care. Physicians, however, must be aware that the optimal threshold for different D-dimer assays may differ across different settings.

References

[1] Anderson J., Wheeler H. B., Goldberg R. J., Hosmer D. W., Patwardhan N. A., Jovanovic B., Forcier A., and Dalen J. E. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. the worcester DVT study, 1991, *Arch Intern Med*, 151(5):933–8.

[2] Ginsberg J. S. Management of venous thromboembolism, 1996, *N Engl J Med*, 335(24):1816–28.

[3] Hull R., Delmore T., Carter C., Hirsh J., Genton E., Gent M., Turpie G., and McLaughlin D. Ad-

justed subcutaneous heparin versus warfarin sodium in the long-term treatment of venous thrombosis, 1982, *N Engl J Med*, 306(4):189–94.

[4] Gardiner C., Pennaneac'h C., Walford C., Machin S. J., and Mackie I. J. An evaluation of rapid d-dimer assays for the exclusion of deep vein thrombosis, 2005, *Br J Haematol*, 128(6):842–8.

[5] Heim S. W., Schectman J. M., Siadaty M. S., and Philbrick J. T. D-dimer testing for deep venous thrombosis: a metaanalysis, 2004, *Clin Chem*, 50(7):1136–47.

[6] Oudega R., Moons K. G., and Hoes A. W. Limited value of patient history and physical examination in diagnosing deep vein thrombosis in primary care, 2005, *Fam Pract*, 22(1):86–91.

[7] De Monye W., Sanson B. J., Buller H. R., Pattynama P. M., and Huisman M. V. The performance of two rapid quantitative d-dimer assays in 287 patients with clinically suspected pulmonary embolism, 2002, *Thromb Res*, 107(6):283–6.

[8] Schutgens R. E., Haas F. J., Gerritsen W. B., Horst F., Nieuwenhuis H. K., and Biesma D. H. The usefulness of five d-dimer assays in the exclusion of deep venous thrombosis, 2003, *J Thromb Haemost*, 1(5):976–81.

[9] van der Graaf F., Borne H., Kolk M., Wild P. J., Janssen G. W., and Uum S. H. Exclusion of deep venous thrombosis with d-dimer testing–comparison of 13 d-dimer methods in 99 outpatients suspected of deep venous thrombosis using venography as reference standard, 2000, *Thromb Haemost*, 83(2): 191–8.

[10] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[11] Perrier A., Desmarais S., Miron M. J., Moerloose P., Lepage R., Slosman D., Didier D., Unger P. F., Patenaude J. V., and Bounameaux H. Non-invasive diagnosis of venous thromboembolism in outpatients, 1999, *Lancet*, 353(9148):190–5.

[12] Schutgens R. E., Ackermark P., Haas F. J., Nieuwenhuis H. K., Peltenburg H. G., Pijlman A. H., Pruijm M., Oltmans R., Kelder J. C., and Biesma D. H. Combination of a normal d-dimer concentration and a non-high pretest clinical probability score is a safe strategy to exclude deep venous thrombosis, 2003, *Circulation*, 107(4):593–7.

[13] Stein P. D., Hull R. D., C. P. K., Olson R. E., Ghali W. A., Brant R., Biel R. K., Bharadia V., and Kalra N. K. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review, 2004, *Ann Intern Med*, 140(8):589–602.

[14] Jaeschke R., Guyatt G. H., and Sackett D. L. Users' guides to the medical literature. iii. how to use an article about a diagnostic test. b. what are the results and will they help me in caring for my patients? the evidence-based medicine working group, 1994, *Jama*, 271(9):703–7.

Chapter 8

The diagnostic accuracy & user-friendliness of five Point of Care D-dimer assays for the exclusion of Deep Vein Thrombosis

The contents of this chapter are based on
Toll DB, Geersing GJ, Oudega R, Wijland R,
Blikman MJC, Moons KGM, Hoes AW
The diagnostic accuracy & user-friendliness of five Point of Care
D-dimer assays for the exclusion of Deep Vein Thrombosis
Submitted

Introduction

More than 75% of the patients suspected of deep vein thrombosis (DVT) who are referred for objective testing -commonly ultrasonography- do not have DVT.[1,2] To improve the diagnostic process and notably to reduce the number of referrals for ultrasonography, D-dimer testing has been introduced. D-dimers are degradation products of cross-linked fibrin generated during fibrinolysis, and can therefore be used as an indirect measure of thrombus formation. Levels of D-dimer are typically elevated in patients with acute DVT, but may also be increased in other conditions such as cancer, pregnancy, and recent surgery. D-dimer assays commonly have good sensitivity, but poor specificity for diagnosing DVT.[3–5] A negative D-dimer test result in combination with a low probability according to a diagnostic decision rule[6,7] is generally considered safe for ruling out DVT.[2,8]

To enable the exclusion of DVT in a near patient situation -without the need for referral for labor intensive and time-consuming laboratory testing- several so-called 'point of care' (POC) D-dimer assays have recently been introduced.[9–15] By means of these assays, a rapid quantitative or qualitative D-dimer test result can be obtained from a patient's whole blood, with the use of hardly any additional laboratory handling. The use of these POC D-dimer assays may especially be useful in settings that have poor accessibility to additional testing (conventional laboratory D-dimer or ultrasonography testing), such as primary in care and nursing homes. However, the diagnostic accuracy of such assays should first be shown in the relevant patient population before using them - in combination with a diagnostic decision rule - in clinical practice. Moreover, the diagnostic accuracy may differ between different POC D-dimer assays, and the results of clinical management studies using conventional (Latex or ELISA based methods) laboratory D-dimer assays cannot necessarily be safely extrapolated to other - notably - POC assays other than those specifically evaluated in these studies.[5,15]

The aim of this study was therefore to directly compare the diagnostic accuracy of five available POC D-dimer assays in primary care patients suspected of DVT. In addition, we assessed the user-friendliness of the five POC assays, because this may assist physicians in determining their preferred assay, particularly when assays show similar diagnostic accuracy.

Patients & Methods

Patients

Between November 2006 and October 2007, we prospectively identified 243 consecutive primary care patients suspected of DVT and referred by their general practitioner

for compression ultrasonography (CUS) to one of the two participating centres (Deventer Hospital or Medical Diagnostic Centre Rotterdam) in the Netherlands. Suspicion of DVT was based on pain, swelling and/or redness of the leg. Exclusion criteria were age below 18 years, any anticoagulant therapy (i.e. vitamin K antagonist or low molecular weight heparin), and inability or unwilling to give informed consent. The study protocol was approved by the medical ethics committees of the two participating centres.

Blood collection

In all patients, venous whole blood was drawn from the anterior cubital vein. Immediately after venous blood collection, capillary blood was collected by means of a fingerprick. Plasma was obtained by centrifuging part of the venous whole blood 3100 $\times$ g for 10 minutes with the ROTIXA 50 RS (Hettich GmbH & Ko KG, Tuttlingen, Germany).

D-dimer assays

The five POC D-dimer assays were performed on each patient's blood sample. All assays were performed using the venous blood sample, except the Clearview®, which uses capillary blood. The assays were performed according to the manufacturer's instructions by technicians unaware of the result of the CUS, and blinded for the outcome of the other D-dimer assays.

Quantitative D-dimer assays

Results of D-dimer measurements are based on assay-specific reference curves. The quantitative test results are either reported as D-dimer units (D-DU, based on calibration with purified fibrin fragment D-dimer), or as fibrinogen equivalent units (FEU, based on the amount of purified fibrinogen used for the preparation of a cross-linked fibrin clot, which is than degraded by plasmin and used as calibrator).[15] A test result given in FEU can be calculated in D-DU, using a conversion factor of 1.8. Consequently, 1 ng/ml FEU corresponds with 1.8 ng/ml D-DU.[15] We studied four quantitative assays.

1. Triage® D-dimer

The Triage® D-dimer assay (Biosite, San Diego, USA) was performed with 250 μl of EDTA whole blood using the Triage® Meter Plus. This single-marker fluorescence

sandwich immunoassay uses the 3B6 D-Dimer antibody. The Triage® Meter Plus displays a quantitative result within a range of 0.1-5 μg/ml D-DU, using a cut-off value of 0.35 μg/ml D-DU.

2. Cardiac D-dimer

The Cardiac D-dimer assay (Roche Diagnostics, Mannheim, Germany) was performed with 150 μl whole blood anticoagulated with Li-heparin, using the Cobas h 232 reader instrument. The assay is based on a dual monoclonal antibody sandwich comprising a poly-(streptadivin)-biotin capture system with a gold particle label. The Cardiac D-dimer displays a quantitative result within a range of 0.1-4.0 μg/ml FEU, using a cut-off value of 0.5 μg/ml FEU.

3. Pathfast^{TM} D-dimer

The Pathfast^{TM} D-dimer assay (Mitsubishi Cagaku Iatron inc, Tokyo, Japan) was performed with 100 μl citrated plasma, using the Pathfast^{TM} in vitro diagnostic system. This chemiluminescent enzyme immunoassay contains two D-dimer monoclonal antibodies, one labelled with alkaline phosphatase and the other coated with magnetic particles. The diagnostic system displays a quantitative result within a range of 0.005-5 μg/ml FEU, using a cut off value of 0.57 μg/ml FEU.

4. Vidas® D-dimer Exclusion^{TM}

The Vidas® D-dimer assay (bioMérieux, Marcy l'Etoile, France) was performed with 200 μl of citrated plasma, using the Vidas® mini analyzer. This assay combines a two-step enzyme immunoassay sandwich method with fluorescent detection. It uses an alkaline-phosphatase labelled anti-FbDP monoclonal mural antibody (P2C5A10). The Vidas® analyzer displays a quantitative result within a range of 45-10.000 ng/ml FEU, using a cut off value of 500 ng/ml FEU.

Qualitative D-dimer assay: Clearview® Simplify® D-dimer

The Clearview® Simplify® D-dimer assay (Inverness Medical, Bedford, UK) was performed with 35 μl of capillary blood. The test principle is based on immunochromatography using two D-dimer specific murine monoclonal antibodies, one of them (DD3B6/224) conjugated to colloidal gold particles. A visible pink-purple coloured line is visible at the test zone (T) when D-dimer levels exceed 80 ng/ml (positive test result). Test results are valid if a pink-purple line is present at the control zone (PC).

Reference standard (compression ultrasonography)

After venous and capillary blood was drawn, each patient underwent real time B-mode CUS with a 5-12 MHz linear-array sonographic scanner (EnVisor HD, Philips Medical Systems, Best, the Netherlands) of the symptomatic leg as a reference standard. DVT was considered present if one of the proximal veins was not fully compressible. The physician who judged the result of the CUS was blinded for the outcome of the D-dimer assays. CUS was repeated within one week in patients displaying a normal ultrasound.[16]

Accuracy of the assays

To determine the overall discriminatory accuracy for in- or exclusion of DVT of the four quantitative D-dimer assays, we calculated the area under the Receiver Operating Characteristic curves (ROC area) with corresponding 95% confidence intervals (CI). Next, the accuracy measures -sensitivity, specificity, negative predictive value- with corresponding 95% CI of the D-dimer assays were calculated at the manufacturers' provided thresholds.

User-friendliness of the point of care D-dimer assays

We also assessed the user-friendliness of the five POC D-dimer assays. Twenty technicians, unfamiliar with D-dimer testing received an elaborate instruction by the manufacturers on the use of the five POC assays and subsequently practiced each assay twice to get acquainted with the measurements. Directly after performing each assay -in a random order- for the third time, the technicians filled in a questionnaire concerning the user-friendliness. The questionnaire included characteristics as: time needed for blood sample collection, time needed for preparation of the assay, time required for delivery of the assay result, difficulty in reading the assay result, liability to flaws in the procedure and the operation friendliness of the analyzer. Time was measured by means of a stop-watch and rounded to minutes. The other questions were measured on a 3-point-scale: small or easy; moderate; large or difficult.

Next to the items included in the questionnaire, two authors (RW and MB) recorded for each assay the time needed for calibration and for applying control assays. Finally, we registered the suitability for home visits, multi marker assay possibility, costs and expiry dates of the required material and devices, the amount of test strips sold in one package, storage temperature, print function, and the maximum amount of simultaneous analyses.

Results

During the study period, 243 primary care patients suspected of DVT were included. Ultrasound measurements were retrievable in 241 (99.2%) of the patients, of which 30 (12%) were indicative of DVT. In 200 patients, both ultrasound results and the results of all five D-dimer assays were available, of whom 24 had DVT. The 43 patients with one or more missing D-dimer assay results or unknown CUS result were excluded from the analysis. The characteristics, age, gender, symptoms, of these 43 excluded patients and the remaining 200 patients were similar (data not shown). The mean age of the 200 analysable patients was 59 years and 67% was female. Additional patient characteristics are presented in Table 8.1.

Table 8.1: Patient characteristics of 200 patients with suspected deep vein thrombosis

Characteristic*	**Value**
Female Gender	67
Mean age in years (SD)	59 (17)
Median days of symptoms (IQR)	9 (5-21)
Previous deep venous thrombosis	16
Females using oral contraception	15
Treatment any malignancy $<$ 6 months	2
Surgery $<$ 4 weeks	2
Leg trauma $<$ 4 weeks	3
Vein distension	3
Calf swelling $\geq$ 3cm	15
Deep vein thrombosis	12

* Values are give in percentage, unless otherwise stated; SD=standard deviation; IQR=inter quartile range

Accuracy measures

The diagnostic accuracy measures with corresponding 95% confidence intervals of the five D-dimer assays are summarized in Table 8.2. The ROC area was 0.87 for the Cardiac and Triage® D-dimer assays and 0.89 for the Vidas® and Pathfast^{TM} D-dimer assays. At the manufacturers' provided thresholds, all D-dimer assays showed negative predictive values higher than 98%. Vidas® assay had a sensitivity of 100%, with a specificity of 40%. Pathfast^{TM}, Triage® and Cardiac D-dimer assays showed a sensitivity of 95.8%, while specificities were 35%, 48%, and 57%, respectively. The only qualitative assay, the Simplify® D-dimer assay, had a sensitivity of 91.7% with a specificity of 63%.

Table 8.2: Diagnostic accuracy of five point of care D-dimer assays for the exclusion of DVT (n=200)

D-dimer assay	Cut-off Value	Sensitivity %(CI)	Specificity % (CI)	NPV % (CI)	ROC area %(CI)
Vidas®	500 ng/ml FEU	100 (93-100)	40.3 (33-48)	100 (97-100)	0.89 (0.83-0.95)
Pathfast^{TM}	0.57 μg/ml FEU	95.8 (88-100)	34.7 (28-42)	98.4 (95-100)	0.89 (0.83-0.96)
Cardiac	0.5 μg/ml FEU	95.8 (88-100)	56.8 (50-64)	99.0 (97-100)	0.87 (0.80-0.94)
Triage®	0.35 μg/ml D-DU	95.8 (88-100)	47.7 (40-55)	98.8 (97-100)	0.87 (0.79-0.95)
Simplify®	N.A.	91.7 (81-100)	63.1 (56-70)	98.2 (96-100)	N.A.

FEU=Fibrin Equivalent Unit; N.A.=Not applicable; D-DU=D-Dimer Unit; CI= 95% Confidence Interval; NPV=Negative Predictive Value; n=Number of patients

User-friendliness

The ease of operation of the D-dimer assays and the liability to making flaws in the procedure were comparable among the different D-dimer assays (Table 8.3). The Simplify® assay differs from the other assays as no analyzer is needed; it can be performed on capillary blood and no calibration is necessary. The time needed to perform a single analysis differed from 10 minutes for the Simplify® assay until 38 minutes for the Vidas® assay. The time needed for calibration (excluding time for preparation of calibration liquids) varied from 0 minutes (no calibration required) for the Simplify® or less than 1 minute for the Cardiac and Triage® assays to 21 and 38 minutes for the Pathfast^{TM} and Vidas® assays, respectively. In addition, calibration of the Vidas® and Pathfast^{TM} assays required some laboratory skills. The retail prices of analyzers varied from $<$5,000 euros (Cardiac and Triage®) to $>$25,000 euros (Pafhfast^{TM}), whereas for the Simplify® D-dimer no analyzer is required.

Discussion

Our comparison of five available POC D-dimer assays that can be applied to exclude DVT in clinically suspected patients revealed high negative predictive values for all assays ($>$98%), while sensitivities varied from 91.7% (Simplify® assay) to 100% (Vidas® assay). The specificity ranges from 35% for the Pathfast^{TM} to 63% for the Simplify® assay. The major differences in the user-friendliness of the assays pertained to (the need for) calibration, time needed to obtain a D-dimer assay result, portability of the device, and the retail prizes.

In many countries, the initial presentation of patients suspected of DVT occurs in primary care. Here, the direct accessibility of additional testing (conventional D-

Table 8.3: User-friendliness of five point of care D-dimer assays

ASPECTS OF USER-FRIENDLINESS	POINT OF CARE D-DIMER ASSAY				
	PATHFAST^{TM}	VIDAS®	SIMPLIFY®	TRIAGE®	CARDIAC
Types of blood samples (anticoagulant)	-venous whole blood -plasma (heparin-Li, heparin-Na, citrate-NA -capillary blood	-plasma (citrate-Na)	-capillary blood -venous whole blood -plasma (citrate-Na, heparin, EDTA)	-venous whole blood -plasma (EDTA)	-venous whole blood (heparin)
Type of blood sampling	vena punction	vena punction	fingerprick or vena punction	vena punction	vena punction
Time (min.) needed for:					
-Warming up analyzer/test strips	30 / 0	30 / 0	NA / 0	$<$1 / 15	$<$1 / 0
-Preparation for analysis	1	1	$<$1	1	1
-Analysis-time	15-17	38	$\leq$10	10-15	10
-Preparation calibration or control liquids	17	24	NA	NA	16
-Control assay / calibration	17 / 21	35 / 44	NA / NA	NA / $<1^a$	20 / $<1^a$
Frequency of:					
-Quality control assay	2 /day	1 / 2 weeks, 1 / lot	NA	NA^b	1 / lot^c
-Control optic system device	NA	NA	NA	NA	c
-Calibration	1 / day, 1 / lot	1 / 2 weeks, 1 / lot	NA	1 / lot^a	1 / lot^a
Calibration by experienced personnel?	yes	yes	NA	no	no
Ease of (% easy / moderate / difficult):					
-Reading test result	95 / 5 / 0	100 / 0 / 0	75 / 25 / 0	95 / 5 / 0	100 / 0 / 0
-Operation of analyzer	95 / 0 / 5	95 / 5 / 0	NA	85 / 15 / 0	90 / 10 / 0
Liability to flaws in procedure (% small / moderate / large):					
-Blood application on teststrip	75 / 20 / 5	60 / 35 / 5	80 / 20 / 0	85 / 15 / 0	90 / 10 / 0
-Buffer application on teststrip	NA	NA	80 / 20 / 0	NA	NA
-Teststrip placement in analyzer	95 / 0 / 5	70 / 20 / 10	NA	100 / 0 / 0	90 / 10 / 0
Simultaneous analyses (amount)	6	12	NA	1	1
Retail price^d					
-analyzer	€€€€	€€	NA (no analyzer)	€	€
-test kit (Euro)	512 / 60 test strips	5-8 / test strip	95 / 10 test strips	250 / 25 test strips	95 / 10 test strips
Max. storage time test strips^{e,f}	12 mo at 2-8 ^{o}C	10-12 mo at 2-8 ^{o}C	21 mo at room temp	8 mo at 2-8 ^{o}C	? at 2-8 ^{o}C
Multiple marker (amount)	7 single markers	84 single markers	no (1 single marker)	7 multi- & 3 single	5 single markers
Print function (yes/no)	yes	yes	no	yes	no
Delayed analysis after blood sample collection (hours at ^{o}C)^f	4h room temp	24h at 2-8 ^{o}C	24h at 2-8 ^{o}C	24h at room temp	8h at room temp

NA: not applicable; a:calibration data transferred by lot specific code chip; b:low and high quality controls are build in the teststrip; c:manufacturer does not recommend a fixed frequency; 1 / lot: once per change of lot number; d:Retail prize in euros (excl VAT) for one analyzer in the Netherlands, 2007; €:$<$ 5,000, €€:5,000-10,000, €€€: 15,000-25,000, €€€€: $>$25,000; e:mo=months; f:temp=temperature; ?:data not provided by manufacturer

dimer and CUS testing) is often limited compared to secondary care. Availability of sensitive POC D-dimer assays is therefore highly valuable in this setting, as well as in elderly homes or nursing home settings. However, to the best of our knowledge, this is the first study exploring and comparing the diagnostic accuracy of POC D-dimer assays in primary care patients. The results of this study can therefore only be compared to similar studies performed in secondary care. Yet, observed differences between primary and secondary care studies may be difficult to interpret, as spectrum bias and variation in DVT prevalence may (partially) explain these dicrepancies.[17]

The results of this study are, however, largely in accordance with the few studies in secondary care that evaluated the accuracy of a POC D-dimer assay for the exclusion of DVT. Concerning the Cardiac D-dimer assay[9,13,15], the reported sensitivity by Bucek et al.[9] was lower (89%) than ours, probably because they also included calf vein thrombosis. In contrast, Legnani et al.[13] showed a relatively high sensitivity (100%), which may be explained by the lower cut-off value applied for a positive test ($>$ 400 ng/ml), compared to $>$ 500 ng/ml (i.e. the cut-off recommended by the manufacturer) in our study. For the Pathfast^{TM} assay, a previous study showed a higher sensitivity (100%) and a much higher specificity (63%) as compared to our study.[12] For the Triage® assay, the accuracy reported by Ghys[18] is in accordance with our results. Regarding the Simplify® assay, reported sensitivities (94% and 100%) are higher and specificities relatively lower (40% and 53%) compared to our results.[10,14] Finally, many studies have evaluated the diagnostic accuracy of the Vidas® D-dimer assay, mostly on the conventional (non-POC) analyzer. Results of these studies are largely in accordance with ours: high sensitivities and mediocre specificities for the exclusion of clinically suspected DVT.[4]

To appreciate the results, a few methodological issues need to be discussed. First, due to the limited amount of patients included in this study, confidence intervals of the accuracy measures were wide, and the accuracy measures should be interpreted with some caution. We recommend additional studies to confirm our findings. Second, to best serve clinical practice, we decided to take the threshold recommended by the manufacturer. Also, larger studies may determine whether other thresholds of the quantitative assays achieve better diagnostic accuracy. Third, in 17% of the patients, one ore more D-dimer test results were not available. These results were missing due to logistic problems (e.g. inability to perform a specific assay due to lack of properly anti-coagulated blood, or temporarily unavailability of particular test strips). Although excluding patients with one or missing D-dimer test results may have limited the power of the study, we found similar characteristics between the 43 excluded patients and the 200 analysed patients. Hence, it is unlikely that the missingness was selective and induced bias. Finally, it is yet still somewhat difficult

to decide which one of the five D-dimer assays should be chosen for daily practice, as it is widely acknowledged that no D-dimer assay - POC or conventional - should be used as a stand-alone test.[4] It is well known that a negative D-dimer test result combined with a low probability according to a diagnostic rule may safely exclude DVT.[6,7] Such combined strategy to "rule-out" DVT has higher sensitivity and negative predictive value than a D-dimer test result alone. The accuracy of the POC D-dimer assays combined with a diagnostic rule was not determined in this study.

Nevertheless, several potentially clinically relevant differences between the five POC D-dimer tests can be observed. As all assays showed very good negative predictive value and similar overall discriminative ability (ROC area), these differences were mainly determined by the sensitivity and specificity at the manufacturer's thresholds and the differences in costs and user-friendliness. The Vidas® assay already showed optimal sensitivity - thus without combination with a diagnostic rule - though at the cost of a mediocre specificity. The sensitivity of the Pathfast^{TM}, Cardiac and Triage® assay was equally good, which was slightly worse compared to the sensitivity of the Vidas® but better compared to the Simplify® assay. However, when combined with a diagnostic rule, the Simplify® assay may achieve the highest exclusion rate of the five POC D-dimer assays. Due to its high specificity, the Cardiac assay outperformed the Pathfast^{TM} and Triage® assays.

This is the first study that evaluated the costs and user-friendliness of the five POC D-dimer assays, making comparisons with the literature impossible. The ease of operation of the D-dimer assays and the liability to making flaws in the procedure were comparable among the different D-dimer assays. All POC assays could easily be performed by technicians without laboratory skills, with a small risk of making flaws in the procedure. However, we found that the user-friendliness of the Vidas® and Pathfast^{TM} assays may be limited in primary care due to the size of the analyzers and because they require relatively difficult and time-consuming calibration. In contrast, the portable and relatively cheap Cardiac and Triage® assays seem feasible for use in primary care, while the Simplify® assay may have the best user-friendliness as no analyzer is needed, it can be performed on capillary blood, and no calibration is necessary. However, due to the lower sensitivity of the Simplify® found in this study, its exclusion ability in combination with clinical characteristics needs to be shown.

In conclusion, all POC D-Dimer assays showed reasonable to good diagnostic accuracy to aid in the diagnosis of DVT. The user-friendliness of the Vidas® and Pathfast^{TM} assays may be limited for use in primary care, in contrast to that of the Simplify®, Cardiac and Triage® assays.

Acknowledgements

We gratefully acknowledge the support by The Netherlands Organization for Scientific Research (ZonMW 945-04-009). We are indebted to the staff of the laboratory of Deventer Hospital and the staff of the Medical Diagnostic Centre Rotterdam for performing the D-dimer analyses for this evaluation. Unrestricted educational grants were received from Unipath Ltd (Inverness Medical), Biosite, Roche, and Mitsubishi. We thank the manufacturers (Biosite, Roche Diagnostics, Mitsubishi Cagaku Iatron inc, bioMérieux, and Inverness Medical) for supplying the D-dimer strips free of charge. The design and conduct of this study, data analysis and manuscript preparation remained independent of these sponsors.

References

[1] Kyrle P. A. and Eichinger S. Deep vein thrombosis, 2005, *Lancet*, 365(9465):1163–74.

[2] Wells P. S., Owen C., Doucette S., Fergusson D., and Tran H. Does this patient have deep vein thrombosis?, 2006, *Jama*, 295(2):199–207.

[3] Goodacre S., Sampson F. C., Sutton A. J., Mason S., and Morris F. Variation in the diagnostic performance of d-dimer for suspected deep vein thrombosis, 2005, *Qjm*, 98(7):513–27.

[4] Heim S. W., Schectman J. M., Siadaty M. S., and Philbrick J. T. D-dimer testing for deep venous thrombosis: a metaanalysis, 2004, *Clin Chem*, 50(7):1136–47.

[5] Kelly J., Rudd A., Lewis R. R., and Hunt B. J. Plasma d-dimers in the diagnosis of venous thromboembolism, 2002, *Arch Intern Med*, 162(7):747–56.

[6] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[7] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[8] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[9] Bucek R. A., Quehenberger P., Feliks I., Handler S., Reiter M., and Minar E. Results of a new rapid d-dimer assay (cardiac d-dimer) in the diagnosis of deep vein thrombosis, 2001, *Thromb Res*, 103(1): 17–23.

[10] Cini M., Legnani C., Cavallaroni K., Bettini F., and Palareti G. A new rapid bedside assay for d-dimer measurement (simplify d-dimer) in the diagnostic work-up for deep vein thrombosis, 2003, *J Thromb Haemost*, 1(12):2681–3.

[11] Dempfle C. E., Suvajac N., Elmas E., and Borggrefe M. Performance evaluation of a new rapid quantitative assay system for measurement of d-dimer in plasma and whole blood: Pathfasttrade mark d-dimer, 2007, *Thromb Res*, 120(4):591–6.

[12] Fukuda T., Kasai H., Kusano T., Shimazu C., Kawasugi K., and Miyazawa Y. A rapid and quantitative d-dimer assay in whole blood and plasma on the point-of-care pathfast analyzer, 2007, *Thromb Res*.

[13] Legnani C., Fariselli S., Cini M., Oca G., Abate C., and 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, 2003, *Thromb Res*, 111(3):149–53.

[14] Neale D., Tovey C., Vali A., Davies S., Myers K., Obiako M., Ramkumar V., and Hafiz A. Evaluation of the simplify d-dimer assay as a screening test for the diagnosis of deep vein thrombosis in an emergency department, 2004, *Emerg Med J*, 21(6):663–6.

[15] Dempfle C. E., Korte W., Schwab M., Zerback R., and Huisman M. V. Sensitivity and specificity of a quantitative point of care d-dimer assay using heparinized whole blood, in patients with clinically suspected deep vein thrombosis, 2006, *Thromb Haemost*, 96(1):79–83.

[16] Fraser J. D. and Anderson D. R. Deep venous thrombosis: recent advances and optimal investigation with us, 1999, *Radiology*, 211(1):9–24.

[17] Knottnerus J. A. Between iatrotropic stimulus and interiatric referral: the domain of primary care research, 2002, *J Clin Epidemiol*, 55(12):1201–6.

[18] Ghys T., Achtergael W., Verschraegen I., Leus B., and Jochmans K. Diagnostic accuracy of the triage((r)) d-dimer test for exclusion of venous thromboembolism in outpatients, 2007, *Thromb Res*.

Chapter 9

General Discussion

The contents of this chapter are based on
Toll DB, Janssen KJM, Vergouwe Y, Moons KGM
Validation, Updating and Impact of Clinical Prediction Rules:
a review
Journal of Clinical Epidemiology, accepted for publication

9.1 Validation, Updating and Impact of Clinical Prediction Rules: a review

Introduction

Prediction rules or prediction models, often also referred to as decision rules or risk scores, combine multiple predictors, such as patient characteristics, test results, and other disease characteristics, to estimate the probability that a certain outcome is present (diagnosis) in an individual or will occur (prognosis). They intend to aid the physician in making medical decisions and in informing patients. Table 9.1 shows an example of a prediction rule.

Table 9.1: Prediction rule for estimating the probability of neurological sequelae or death after bacterial meningitis.[1] The relative weights of the predictors are expressed with regression coefficients and with simplified scores.

	Regression coefficient[a]	**Score**[b]
Male gender	1.48	2
Atypical convulsions	2.27	3
Body temperature -35^{o}C	-0.75	-1
Pathogen		
S. Pneumoniae	3.12	4
Neisseria meningitidis	1.48	2
Intercept / constant	25.2	5

Score	**$<$ 2.5**	**2.5 - 4.5**	**5.0 - 5.5**	**$>$ 5.5**	**Overall**
Probability	0/25 (0%)	2/78 (2.6%)	6/33 (18.2%)	15/34 (44.1%)	23/170 (13.5%)

[a] The probability for each patient can be calculated as $log(\frac{\text{risk of outcome}}{1-\text{risk of outcome}}) = 25.2 + 1.48 \times$ male gender $+ 2.27 \times$ atypical convulsions $- 0.75 \times$ (body temperature $- 35^oC) + 3.12 \times$ S. Pneumoniae $+ 1.48 \times$ N. Meningitidis

[b] The scores were derived by dividing the regression coefficients of the included predictors by the smallest regression coefficient and then rounding them to the nearest integer. For each patient, a sumscore can be calculated by adding the scores that correspond to the characteristics of the patient. The total sumscore are related to the individual probability as shown in the lower part of Table 9.1.

In multivariable prediction research, three phases may be distinguished; 1. development of the prediction rule; 2. external validation of the prediction rule (further referred to as 'validation'), i.e. testing the rule's accuracy and thus generalisability in data that was not used for the development of the rule, and subsequent updating if validity is disappointing; and 3. studying the clinical impact of a rule on physician's behaviour and patient outcome (Table 9.2).[2–6] A quick Medline-search using a sug-

gested search strategy[7] demonstrated that the number of scientific articles discussing prediction rules has more than doubled in the last decade; 6,744 published articles in 1995 compared to 15,662 in 2005. A striking fact is that this mainly includes papers concerning the development of prediction rules. A relatively small number regards the validation of rules and there are hardly any publications showing whether an implemented rule has impact on physicians behaviour or patient outcome.[4,5]

Table 9.2: Consecutive phases in multivariable prediction research.

Phase	Short description
1 Development	Development of a multivariable prediction rule, including identification of important predictors, assigning the relative weights to each predictor, estimating the rule's predictive accuracy, estimating the rule's potential for optimism using so-called internal validation techniques, and -if necessary- adjusting the rule for overfitting.
2 Validation and updating	Testing the accuracy of the prediction rule in patients that were not included in the development study. Temporal, geographical, and domain validation can be distinguished. If necessary, the prediction rule can be updated, by combining the information captured in the rule (development study) and the data of the new patients (validation study).
3 Impact	Determining whether a (validated) prediction rule is used by physicians, changes therapeutic decisions, improves patient outcome or will reduce costs.

Lack of validation and impact studies is unfortunate, because accurate predictions - commonly expressed in good calibration (agreement between predicted probabilities and observed outcome frequencies) and good discrimination (ability to distinguish between patients with and without the outcome)- in the patients that were used to develop a rule are no guarantee for good predictions in new patients, let alone for their use by physicians.[2,4,5,8,9] In fact, most prediction rules commonly show a reduced accuracy when validated in new patients.[2,4,5,8,9] There may be two main reasons for this: 1. the rule was inadequately developed; or 2. there were (major) differences between the derivation and validation population.

Many guidelines regarding the development of prediction rules have been published, including the number of potential predictors in relation to the number of patients, methods for predictor selection, how to assign the weights per predictor, how to shrink the regression coefficients to prevent overfitting, and how to estimate the rule's potential for optimism using so-called internal validation techniques such as bootstrapping.[2,3,8–15]

Compared to the literature on the development of prediction rules, the methodology for validation and studying the impact of prediction rules is underappreciated.[2,5,9]

This paper provides a short overview of the types of validation studies, of possible methods to improve or update a previously developed rule in case of disappointing accuracy in a validation study, and of important aspects of impact studies and implementation of prediction rules. We focus on prediction rules developed by logistic regression analysis, but the issues largely apply to prediction rules developed by other methods such as Cox proportional hazard analysis or neural networks.

Examples of disappointing accuracy of prediction rules

Even when internal validation techniques are applied to correct for overfitting and optimism, the accuracy of prediction rules can be substantially lower in new patients compared to the accuracy found in the patients of the development population. For example, the generalisability of an internally validated prediction rule for diagnosing a serious bacterial infection in children presenting with fever without apparent source was disappointing.[16] In the development study, the area under the receiver operating characteristic curve (ROC area) -after adjustment for optimism- was 0.76 (95% confidence interval [CI]: 0.66-0.86). However, when applied to new patients obtained from another hospital in a later period using the same in- and exclusion criteria, the ROC area dropped to 0.57 (95% CI: 0.47-0.67). The authors concluded that this could partly have been caused by flaws in the development of the rule, notably too few patients in relation to the number of predictors tested, but also that internal validation and correction for optimism do not always prevent a decreased accuracy in future patients.[16]

Another example of poor generalisability regards the European System for Cardiac Operative Risk Evaluation (EuroSCORE), a prediction rule that was developed in 128 centres in eight European states to predict 30-day mortality in patients who underwent cardiac surgery.[17,18] Validation studies showed good results in European, North American and Japanese populations.[17,19–24] Yap et al.[25] tested the generalisability of the EuroSCORE in 8331 cardiac surgery patients from six Australian institutions and found that predictions were poorly calibrated for Australian patients. According to the authors, reasons for this finding are unclear and likely to be multi-factorial, such as different health care system, different indications for cardiac surgery, and different prevalence of co-morbid conditions in Australia compared to Europe. Also, the prediction rule could be 'out of date'.[25,26] The rule was developed with data of patients who were operated more than 10 years prior to the patients in the Australian validation study. The surgical procedure has indeed changed over time, potentially leading to different outcomes.[25,26] This change was not reflected by the other validation studies.[17,19–24]

Common differences between a development and a validation population

As described, disappointing generalisability can be explained by differences in the development and validation population. We may largely identify three possible differences. First, the definitions of predictors and the outcome variable, and the measurement methods may be different.[2,3,9,12] Prediction rules that contain unclear defined predictors or predictors which measurement or interpretation is liable to subjectivity are likely to show a reduced predictive strength when applied to new patients. For example, the prediction rule of Table 9.1 contains the rather objective predictor gender, but also the presence of 'atypical convulsions'.[1] The latter may be defined differently by physicians, which may compromise the generalisability of the rule. It is advised to determine the interobserver variability of potential predictors, and to include only those predictors in the final prediction rule that show good reliability.[3,6] Improvement in measurement techniques for predictors may also affect the predictive strength of a predictor. For example, the Magnetic Resonance Imaging (MRI) technique is developing rapidly over time, which results in improved image quality. Consequently, the diagnostic or prognostic information of MRI probably also improves over time and influences the accuracy of prediction rules that include MRI information.

Second, the group of patients used for the development of a prediction rule may be different from the group of patients used for validation. This is also called difference in 'case-mix'.[2,27] For example, differences in indication for cardiac surgery and differences in co-morbidity were considered as one of the causes of the poor calibration of the (prognostic) EuroSCORE in Australian patients. Both discrimination and calibration of a rule can be affected by differences in case-mix. For example, a validation population may only include elderly (e.g. defined as age $\geq$ 65 years), while in the development population individuals' age ranged from 18–85 years. If age is a predictor in the rule, then discrimination between presence or absence of the outcome in the more homogeneous validation population is more difficult than in the more heterogeneous development population. Further, a validation population may e.g. contain relatively more males than the development population. If male gender increases the probability of the outcome but gender was not included in the rule (missed predictor), then the predicted probabilities by the rule will be underestimated in the validation population (reduced calibration).[27]

Third, validation studies commonly include fewer individuals than development studies. Accordingly, both populations may seem different, which is notably due to random variation.[13,28] The required size of a validation study depends on the hypotheses tested. For prediction rules that predict dichotomous outcomes, it has been suggested that the validation sample should contain at least 100 events and 100 non-

events to detect substantial changes in accuracy with 80% power, e.g. a 0.1 change in c-statistic.[29]

Type of validation studies

It has repeatedly been suggested that a validation study should consist of an adequate sample of 'different but related patients' compared to the development study population.[2,5,9] Relatedness is at least defined as 'patients suspected of the same disease' for a diagnostic rule, and for a prognostic rule as 'patients at risk of the same event'.

In hierarchy of increasingly stringent validation strategies, we largely distinguish temporal, geographical, and domain validation.[2,4,5,9] In general, the potential for differences between the development and validation population is smallest in a temporal validation study, and largest in a domain validation study (Table 9.3). Consequently, confirmative results in a domain validation study are considered to provide the strongest evidence that the prediction rule can be generalised to new patients, while the generalisability of a prediction rule which has shown confirmative results in a temporal validation study may still be restricted.

Temporal validation

Temporal validation tests the generalisability of a prediction rule 'over time'. In a temporal validation study, the prediction rule is typically tested by the same physicians or investigators as in the development study, in the same institution(s), and in similar patients e.g. using the same eligibility criteria resulting in small variation in case-mix (Table 9.3).[4,5,9] Hence, this type of validation is usually successful for thoughtfully developed prediction rules. However, improvements in medical techniques may still affect the predictive accuracy of a prediction rule, as in the example of the EuroSCORE. Confirmative results in (multiple) temporal validation studies indicate that clinicians may cautiously use the prediction rule in their future patients who are similar to the development and validation population. But validation in varied study sites is still necessary before the rule can be implemented in other (geographical) locations or other patient domains (see below).[2,5,9]

Geographical validation

Geographical validation studies typically test the generalisability of a prediction rule in a patient population that is similarly defined as the development population (as is the case in temporal validation studies), though in hospitals or institutions of other geographical areas.[2,9] 'Other geographical areas' can be within one country, across similar countries (e.g. western to western or non-western to non-western countries),

Table 9.3: Potential differences between the development and validation population and the influence on generalisability of the prediction rule: - weak; +/- possible; + probable; ++ likely.

	Temporal validation:	**Geographical validation:**	**Domain Validation:**
	To test the generalisability of a prediction rule 'over time' in similar patients as in the development study (same in- and exclusion criteria) in the same hospitals or institutions.	To test the generalisability of a prediction rule in similar patients as in the development study (same in- and exclusion criteria) in hospitals or institutions of another geographical area.	To test the generalisability of a prediction rule across different domains, which may contain other patient (sub)groups.
Differences in:			
Interpretation of predictors and outcome	- Often same physicians as in the development study who are thus experienced in obtaining the predictors and outcome; differences in interpretation of predictors is less unlikely	+ Other physicians as in the development study, who may define subjective predictors (and the outcome) differently; differences in interpretation of predictors may occur	+ Other physicians as in the development study, who may define subjective predictors (and the outcome) differently; differences in interpretation of predictors may occur
Used measurements for predictors and outcome	+/- Same measurements used, unless measurements have been replaced (time-related)	+ Other measurements may be used ('institution dependent'), plus potential for time-related changes in measurements	+ Other measurements may be used ('institution dependent'), plus potential for time-related changes in measurements
Case-mix	- Patient populations are similar; (random) variation due to a commonly small sample size of the validation population is possible	+/- (Subtle) differences in case-mix possible, beyond possible (random) variation due to a commonly small sample size of the validation population	++ Differences in case-mix are (very) likely, beyond possible (random) variation due to a commonly small sample size of the validation population

and across non-similar countries (e.g. western to non-western or vice versa). Understandably, the less similar the development and validation locations are, the more potential for differences in 'interpretation of predictors and outcome', 'measurements used' and 'case-mix', and thus the more potential for disappointing generalisability (Table 9.3).

Physicians participating in a geographical validation study may be less experienced using the prediction rule, which may influence the accuracy of the rule, due to predictors sensitive to subjectivity. Further, measurement of predictors and the outcome can be performed with different methods than in the development study,

also potentially affecting the rule's accuracy. For example, prediction rules to safely exclude the diagnosis deep vein thrombosis (DVT) contain items from patient history, physical examination, and the result of a D-dimer test.[30,31] Many different D-dimer assays are available, each with different diagnostic accuracy measures. Sensitivities of D-dimer tests for diagnosing DVT may vary between 48% and 100%, and specificities from 5% to 100%.[32] This variation in the accuracy of different D-dimer tests will obviously be reflected in the accuracy of the diagnostic rules that include D-dimer tests. Further, if different cut-off values are used to dichotomise a continuous variable to define a positive versus negative result, the predictive accuracy of the same variable may be different across studies. For instance, if an abnormal D-dimer concentration has been defined as higher than 500 ng/ml in the development sample and as higher than 1,000 ng/ml in the validation sample, the rule will likely perform differently in the validation sample.

Case-mix differences in a geographical validation study can be subtle. For example, a rule containing 'antibiotic use in previous month' as a predictor to estimate the probability of 30-day hospitalisation or death from lower respiratory tract infection in elderly patients[33] may show decreased accuracy in another country, not because the predictor was not well defined or liable to subjectivity, but because the indication for prescribing antibiotics, and thus the characteristics of antibiotic receivers, may vary between countries.

Domain validation

Perhaps the broadest form of validation is to test the generalisability of a prediction rule across different domains, such as patients from a different setting (primary, secondary or tertiary care), inpatients versus outpatients, patients of different age categories (e.g. adults versus adolescents or children), of a different gender, and perhaps from a different type of hospital (academic versus general hospital). Obviously, the case-mix of a patient population of a new domain will differ from the development population (Table 9.3), which is usually reflected in differences in the distribution of the predictor values and in the ranges of predictor values.

For example, the case-mix of primary care and secondary care patients is often clearly different. Primary care physicians always selectively refer patients to specialists. These referred patients commonly have relatively more severe signs or symptoms, or have a relatively more developed disease stage.[9,34] Consequently, secondary care patients commonly have a narrower range of the (more severe) predictor values than primary care patients. To some extent, a secondary care population can be considered as a subdomain of the primary care population. Hence, in contrast, validating a prediction rule developed from secondary care in a more heterogeneous primary

care population actually concerns the estimation of the rule's ability for extrapolation. Extrapolation of prediction rules developed in secondary care to primary care patients often results in a decreased accuracy.[9,34] For example, it has been shown that a prediction rule for safely excluding the diagnosis DVT developed in secondary care[31,35] showed disappointing accuracy in primary care: 0.9% of the secondary care patients had DVT while the diagnosis was ruled out according the rule, whereas this proportion was increased to 2.9% in primary care patients.[36]

We note, however, that certain inclusion criteria in development studies may have been chosen for practical reasons only, and may not compromise the generalisation of a prediction rule derived from such studies. For example, the Ottawa ankle rule for safely excluding fractures without additional *x* ray testing[37] was developed on patients aged 18 years or older. One could question whether this rule -which does not include age as a predictor- is accurate when applied to children or adolescents. If the relative weights (odds ratios) of the predictors in the rule are independent of age, extrapolation of the rule to adolescents can be as successful as in the development population. A recent review concerning the extrapolation of the Ottawa ankle rule to children or adolescents indeed concluded that 'a small' percentage (1.4%) of patients that are excluded from receiving *x* ray evaluation based on the Ottawa ankle rule will actually have a fracture,[38] compared to 0% in the development study (in the development study, the score threshold was specifically chosen to achieve a 100% negative predictive value).

Updating prediction rules

When a validation study shows disappointing results, researchers are often tempted to reject the rule and directly pursue to develop new rules with the data of the validation population only. However, while the original prediction rules usually have been developed with large datasets, validation studies are frequently conducted with much smaller patient samples. The redeveloped rules are thus also based on smaller samples. Furthermore, it would lead to many prediction rules for the same outcome, obviously creating impractical situations as physicians have to decide on which rule to use. For example, there are over 60 published rules to predict outcome after breast cancer.[39] Moreover, when every new patient sample would lead to a new prediction rule, prior information that is captured in previous studies and prediction rules would be neglected. This is counterintuitive to the intention that scientific inferences should be based on data of as many patients as possible. This principle of using prior knowledge from previous studies has been recognized and utilized in etiologic and intervention research, for example in the realm of (cumulative) meta-analyses.

A logical alternative to re-developing prediction rules in each new patient sample, is to update existing prediction rules with the data of the new patients in the validation study. As a result, updated rules combine the prior information that is captured in the original rules with the information of the new patients of the validation population.[40–43] Hence, updated rules are adjusted to the characteristics of the new patients, and likely show improved generalisibility.

Several updating methods have been proposed in the literature.[40–43] The methods vary in extensiveness, which is reflected by the number of parameters that is adjusted or re-estimated (Table 9.4). We will briefly describe these methods and refer to the literature for a more profound description.[40–43] In many situations, as described before, differences in outcome incidences are found between the development data and the validation data. For example, in a primary care setting one can validate a secondary care rule that predicts the presence or absence of DVT. Due to the higher prevalence of DVT in secondary care,[34] the calibration of the rule in primary care patients may be poor as a result of systematically too high predicted probabilities. By adjusting only the intercept of the original prediction rule for the patients in the primary care setting, the poor calibration can be improved.[42,43] This method is by far the simplest updating method as only one parameter of the original rule, i.e. the intercept, is adjusted (Table 9.4, method 1).

Table 9.4: Updating methods for prediction rules.

No.	Updating method	Reason for updating
0	No adjustment (the original prediction rule)	-
1	Adjustment of the intercept	Difference in outcome incidence
2	Adjustment of the regression coefficients of the predictors (by one adjustment factor) and of the intercept	Regression coefficients of the original rule are overfitted
3	Method 2 + extra adjustment of regression coefficients for predictors with a different strength in the validation population compared to the development population	As in method 2, and the strength (regression coefficient) of one or more predictors may be different in the validation population
4	Method 2 + stepwise selection of additional predictors	As in method 2, and one or more potential predictors were not included in the original rule
5	Re-estimation of all regression coefficients, using the data of the validation population	The strength of all predictors may be different in the validation population
6	Model 5 + stepwise selection of additional predictors	As in method 5, and one or more potential predictors were not included in the original rule

Another updating method is called 'logistic recalibration' and can be used when the regression coefficients (relative weights that represent the predictive strength) of the predictors in the prediction rule are overfitted in the development study.[42,43] This typically results when too many predictors were considered in a too small dataset.[8,14] In the lower range, the predicted probabilities in the new patients are usually too low, while in the higher range they are too high. When overfitting was not adequately prevented or adjusted during development of the rule, all regression coefficients can still be adjusted with a single correction factor that is easily estimated from the data of the new patients in the validation set (Table 9.4, method 2).

Although calibration can indeed be improved by these first two methods, discrimination (ROC area) will remain unchanged, as the relative ranking of the predicted probabilities remain the same. To improve the discrimination of a rule in new patients, more rigorous adjustments need to be made to the prediction rule, also called model revisions. We will briefly explain four revision methods that can also improve the discrimination of a prediction rule when a rule is validated in new patients.

First, the strength of one or more predictors can be different in the validation population compared to the development population, while the relative sizes of the other regression coefficients to each other are correct. The regression coefficients that differ can be re-estimated from the validation data (Table 9.4, method 3). For example, we discussed a prediction rule with antibiotic use as a predictor (subsection 9.1). When this rule is applied in a population or setting with a different antibiotic prescription strategy, the strength of this particular predictor may be different, while the strength of the other predictors in the rule not necessarily changes.

Also, when potential predictors that may have predictive value were not included in the original rule, one can test whether these have added predictive value in the validation data (Table 9.4, method 4). For example, when a prediction rule is validated over time, and a new test has become available, the new test may have added predictive value in the rule.

Finally, when the previous described updating methods can not improve the accuracy of the rule, and the strengths of all predictor are expected to be different in the new patients, the intercept and the regression coefficients of all predictors can be re-estimated with the data of the new patients (Table 9.4, method 5). If necessary, additional predictors can be considered as well (Table 9.4, method 6). These two methods are the most rigorous updating methods, as the intercept, regression coefficients and, possibly also additional predictors, are all re-estimated from the validation set. Both methods will probably be most applicable to domain validation, as these are typical situations in which the strength of predictors may differ between the two populations. Note that a disadvantage of these rigorous updating methods is

that the rule is redeveloped on the data of the validation set only and that the prior information in the original rule is neglected, as we discussed above.

With all above described methods, the updated rules are adjusted to the circumstances of the validation population. However, we recommend that updated prediction rules, just like newly developed rules, still need to be tested on their generalisability and impact before they can be applied in daily practice. Note that for all updating methods, data of the new patients is needed. When this data is not available, but one knows the incidence of the outcome and the mean value of the predictors in the new population, the rule can be adjusted by a simple adjustment of the prediction rule.[44,45]

Impact analysis

To ascertain whether a validated prediction rule will actually be used by physicians, will change or direct physicians' decisions, and will improve patient outcomes or reduces costs, an impact study or impact analysis should be performed.[4,5] In the ideal design of an impact study, physicians or care units are randomized to either the index group - which is 'exposed' to the use of the prediction rule - or to the control group using 'care or clinical judgment as usual'.[4] Randomization of patients instead of physicians -such that a physician randomly uses the prediction rule or applies 'usual care'- is not advised. Learning effects will lead to a reduced contrast between the two study groups, resulting in a diluted measured impact of the rule. Moreover, randomising centres (requiring a multi-centre study) instead of physicians within a single centre may prevent the risk of contamination - i.e. exchange of experiences and information by physicians between the two study groups - also leading to reduced contrast and dilution of the rule's effect. Patients are followed up to determine the impact of the prediction rule on patient outcome and on cost-effectiveness. Follow-up is not required for studying the influence of a prediction rule on decision making behavior: a randomized study then suffices. An alternative design to determine the impact of a prediction rule is a before-after study within the same physicians or care units,[46] although temporal changes may compromise the validity of this design.

Although an impact analysis is the method par excellence to study the real effect of a prediction rule in practice, only a limited number of impact analyses have been performed.[5] One of these few studies regarded the impact of a prediction rule aimed at improving the effectiveness of treatment of patients presenting with community-acquired pneumonia to the emergency department, measured by the health-related quality of life and the number of bed days per patient.[47] Hospitals were randomly assigned to implementation of the prediction rule or conventional care, by using a

computer that stratified for type of institution (teaching or community hospital) and average historical length of stay. Physicians were instructed to use the prediction rule as a guide only; the rule did not supersede clinical judgment. An educational plan was designed to reinforce compliance with the use of the prediction rule. The authors concluded that the prediction rule did not improve the health-related quality of life of the patients but reduced the number of bed days per patient managed. This effect was never revealed if no impact study had been conducted.

Implementation of prediction rules

When a rule has frequently been proven to be accurate in diverse populations, the more likely it is that the prediction rule can be successfully applied in practice.[2,5,9] Yet, there are still reasons why the rule is not as successful in daily practice.

First, physicians may feel that their often implicit estimation of a particular predicted probability is at least as good as the probability calculated with a prediction rule, and may therefore not use or follow the rule's predictions.[4] Or, the physicians' estimation of a probability may even have proven to outperform the discrimination of a prediction rule. In a recent review, Sinuff et al. compared the discrimination of physicians' estimations with prediction rules in predicting the mortality of critically ill patients considered for intensive care unit (ICU) admission.[48] They concluded that ICU physicians more accurately discriminate between survivors and non-survivors in the first 24 hours of ICU admission than prediction rules do. It may thus be important to compare physicians' predictions with those of prediction rules, preferably already during the development phase of a prediction rule (requiring obviously a prospective design), but certainly in validation and impact studies. If physicians' predictions of probabilities have proven to outperform the probabilities provided by a rule, the rule will likely not be used in practice. In contrast, results in favour for the prediction rule can be used to convince physicians of using the rule when properly validated.

Second, prediction rules must have face validity; physicians must accept the logic, as well as the science of the rule. Clinical prediction rules that do not have face validity may not be applied in practice, even when effective.[49]

Third, prediction rules may not be used because they are not user-friendly.[4,50] The user-friendliness should be taken into account when developing the rule. A variable should only be considered as a potential predictor if obtaining this variable is also feasible in daily practice of the type of patients under study (not too time-consuming or costly). The user-friendliness of a prediction rule also depends on the way a prediction rule is presented; the original regression formula (as e.g. in the legend of Table 1) is the most exact and accurate form, but may involve cumbersome calculations

requiring a calculator or computer. Although sumscores, risk stratification charts or nomograms may be less precise, they certainly are more user-friendly. With the introduction of electronic patient records, the use of regression formulas in daily practice will become much easier.

Finally, practical barriers may exist to act on the results of the prediction rule. For instance, when using a diagnostic prediction rule aiming to determine whether subsequent testing is necessary, such as the Ottawa ankle rules, physicians may be concerned about protecting themselves against litigation. Hence, they may still refer their patients for additional testing, while at the same time the prediction rule indicates that referral was not necessary.[4] Brehaut et al.[51] conducted a postal survey among 399 randomly selected physicians to examine whether physicians used the Ottawa ankle rules[37] for diagnosing ankle fractures. Most physicians (90%) reported to use the Ottawa ankle rules always or most of the time in appropriate circumstances, while only 42% actually based their decisions to order radiography primarily on the rule.[51] The same authors assessed why some prediction rules become widely used while others do not,[52] taking the Canadian Cervical Spine Rule[53] as an example. They showed that older physicians and part-time working physicians were less likely to be familiar with the rule. The best predictors whether a rule would be used in practice were the familiarity acquired during training, the confidence in the usefulness of the rule, and the user-friendliness of the rule.[52]

Final comments

We have given an overview of types of validation studies, of methods to improve or update a previously developed rule in case of disappointing accuracy in a validation study, and of aspects of impact studies and the implementation of prediction rules. A validated, and if necessary updated, rule may cautiously be applied in new patients that are similar to the patients in the development and validation populations. However, when the user has reasons to believe that the rule may perform differently in the new patients, data of the new patients should first be collected to test the accuracy, and preferably impact of the rule, before it rule is applied in daily clinical practice. Any rule may perform slightly different in a new patient sample due to sampling variation. In that situation, the rule does not need to be updated. The questions remains: when has a rule been sufficiently validated and updated? So far, this particular methodological area of prediction research has not been explored. Future research should address the question how many validation studies and what type of adjustments are needed before it is justified to implement a prediction rule into clinical practice.

Another subject of prediction research that may need more focus in the future is the methodology for systematic reviews and perhaps even meta-analyses of prediction rules for the same outcome.[54,55] It will be a challenge to define how regression coefficients of prediction rules can be combined, and how to properly address publication bias; as prediction rules with good results are more likely to be published than rules with moderate results. Although the methodology for meta-analyses has been extensively described for etiologic and intervention studies, to our knowledge no research has been conducted for meta-analysis to combine several prediction rules.

Last, the potential gain in predictive accuracy and generalisability of a prediction rule developed on combined datasets with individual patient data from various studies on the same outcome (so-called individual patient studies) is a research area that needs more attention.[56]

Our purpose was to stress the importance of testing the generalisability and impact of prediction rules, and outline the methods of such research. The relevance and importance of validating and testing the impact of prediction rules on physician's behaviour, and patients' outcome, has repeatedly been emphasized in the literature. Unfortunately, only a relatively small number of rules are validated, and hardly any study questions whether an implemented rule can change patient outcome. An increased focus on validation and impact studies will likely improve the application of valid prediction rules in daily clinical practice.

9.2 Consecutive phases in multivariable prediction research illustrated by the diagnostic rule for excluding DVT in primary care

In this final section, we briefly illustrate the consecutive phases in multivariable prediction research as described in chapter 9.1, with the diagnostic rule that has been described in chapters 1 to 6. This rule was developed for primary care to safely exclude deep vein thrombosis (DVT) and to reduce the number of unnecessary referrals to secondary care. The process is schematically depicted in figure 9.1.

Clinical problem

In many countries, the primary care physician is the first to encounter patients with symptoms and signs suggestive of DVT. They are challenged with the difficult task to discriminate patients with DVT from those without DVT. The risk of missing the diagnosis and the risk of unnecessary referral or treatment with a potential harmful therapy has to be balanced. The American Academy of Family Physicians and the American College of Physicians recommend that prediction rules to aid in diagnosing DVT should be applied in practice.[57,58] Such a rule for safely excluding DVT without ultrasound testing is available (validated and implemented) in secondary care.[31]

Our first study included a validation of this secondary care DVT rule in primary care patients. We found that the rule could not adequately exclude DVT in primary care patients.[36] Formal updating of the rule - as described in the previous section - could not be realized, because no underlying mathematical formula of the rule was available. This formula was lacking since the rule was assembled from information obtained by literature review and from the collective experience of the participating investigators, and not by using statistical modelling.

Rule development

In contrast to many validation studies, the data set collected for the validation of the secondary care DVT rule was rather large (N=1295 primary care patients suspected of DVT). This allowed us to develop (using multivariable logistic regression modelling) and internally validate a particular prediction rule for primary care, aimed at safely excluding DVT and reducing the number of unnecessary referrals for further work-up in secondary care (Figure 9.1, Phase 1).[30] This new rule could very well discriminate the 'very low' risk patients (who do not need referral to secondary care) from patients

with an increased risk of DVT. The data showed that application of the rule may reduce the number of patient referrals substantially (23%), at the cost of an acceptable low proportion of DVT in non-referred patients (0.7%). The question remained, what studies were required to confirm this good accuracy of the rule - as developed - in new patients to allow for use of the rule in routine primary care?

Subsequent studies

Confirmation in at least one (temporal) validation study - in a similar setting and among similar patients - is required. We thus validated the accuracy of the diagnostic rule in a new cohort of primary care patients suspected of DVT, included by the same physicians as in the development study (Figure 9.1, Phase 2).[59] Additionally, we tested whether the performance of the rule varied across clinically relevant patient subgroups, such as the elderly or patients with a history of DVT and/or pulmonary embolism (PE). Knowledge of these potential modifying factors would enhance the implementation of the diagnostic rule.[60,61] In both studies, the rule was applied in all patients and subsequently they all underwent ultrasound testing (reference test) irrespective of the result of the rule.

Both studies showed that the rule can be used to safely exclude DVT in primary care patients suspected of the disease, also in the patient subgroups. The rule did not need to be updated based on these validation studies.

Promising results on such validation studies increase the confidence to apply the rule in daily practice. However, to formally ascertain whether implementation of the rule in primary care will actually improve the diagnostic process, i.e. a cost-effective reduction of the number of patient referrals without too many false negative diagnoses, an impact or management study was conducted (Figure 9.1, Phase 3). General practitioners actually used the diagnostic rule to decide whether a patient clinically suspected of DVT should be referred for additional testing or could remain under their own supervision. In contrast to the previous (validation) studies, patients designated by the rule as having a very low risk of DVT were now indeed not referred for ultrasonography. The results showed that the rule is not only safe (the number of patient referrals for ultrasound measurements was reduced by almost 50%, at the cost of just 1.4% subsequent venous thromboembolic events in non-referred patients), but also cost-effective as compared to hospital based strategies.

With the data of this impact study, we determined whether the rule could be further optimized by using updating methods (described in the previous section). We

tested whether the strength of the individual predictors of the rule was similar in the new data, and whether inclusion of additional predictors could improve the original rule. We found that further optimization of the rule - to increase the proportion of patients that do not need referral and decrease the number of DVT cases among non-referred patients - was not necessary.

These results illustrate that the diagnostic rule as developed in 2005 by Oudega and colleagues[30] remains the optimal strategy for excluding DVT in primary care.

Remaining issues

(National) implementation of the rule in primary care needs to be ensured (Figure 9.1, Phase 4). To further facilitate the application of the diagnostic rule in primary care practice, a few issues need to be considered. First, physicians should become familiar with the rule. Currently, the Dutch general practitioners guideline ('NHG standaard') "Deep Vein Thrombosis" is being formulated based on the results of the above mentioned studies. A handout containing the diagnostic rule and a short summary of the guideline may more easily be stored within reach and (re)read than the complete guideline. This potentially increases familiarity and use of the rule. Besides, workshops dedicated to the guideline for diagnosing DVT in primary care are organized to improve the use of rule by general practitioners. Second, because implementation of the diagnostic rule may replace part of the diagnostic process from secondary care to primary care, a reimbursement should be established to compensate general practitioners for the extra time and costs accompanied with the use of the rule. Finally, an application for PDAs or for the electronic patient record containing the diagnostic rule may assist in the availability and use of the rule in practice. Possibly, the reimbursement for the use of the rule by general practitioners can be linked with the application in PDAs.

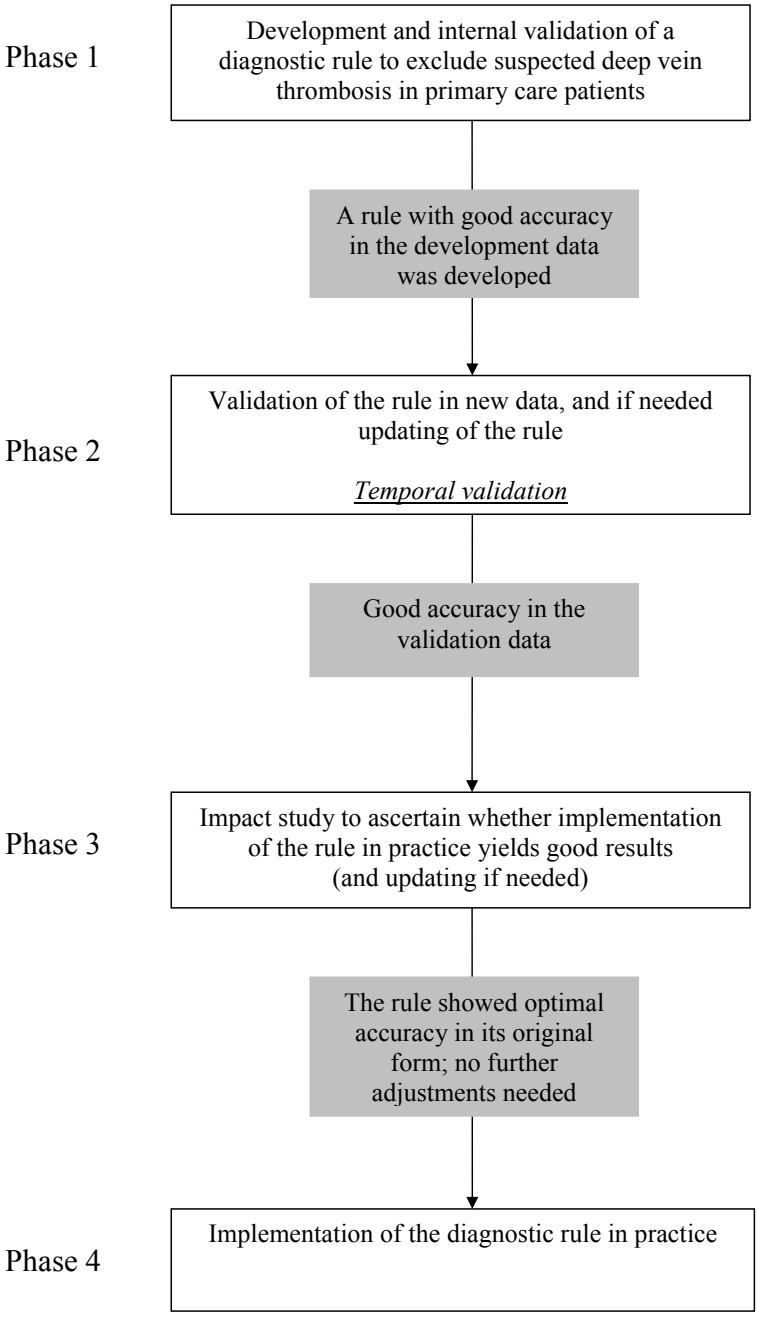

Figure 9.1: Executed consecutive phases in the development, validation, impact and application of a diagnostic rule for clinically suspected deep vein thrombosis in primary care.

References

[1] Oostenbrink R., Moons K. G., Derksen-Lubsen G., Grobbee D. E., and Moll H. A. Early prediction of neurological sequelae or death after bacterial meningitis, 2002, *Acta Paediatr*, 91(4):391–8.

[2] Altman D. G. and Royston P. What do we mean by validating a prognostic model?, 2000, *Stat Med*, 19(4):453–73.

[3] Laupacis A., Sekar N., and Stiell I. G. Clinical prediction rules. a review and suggested modifications of methodological standards, 1997, *Jama*, 277(6):488–94.

[4] McGinn T. G., Guyatt G. H., Wyer P. C., Naylor C. D., Stiell I. G., and Richardson W. S. Users' guides to the medical literature: Xxii: how to use articles about clinical decision rules. evidence-based medicine working group, 2000, *Jama*, 284(1):79–84.

[5] Reilly B. M. and Evans A. T. Translating clinical research into clinical practice: impact of using prediction rules to make decisions, 2006, *Ann Intern Med*, 144(3):201–9.

[6] Wasson J. H., Sox H. C., Neff R. K., and Goldman L. Clinical prediction rules. applications and methodological standards, 1985, *N Engl J Med*, 313(13):793–9.

[7] Ingui B. J. and Rogers M. A. Searching for clinical prediction rules in medline, 2001, *J Am Med Inform Assoc*, 8(4):391–7.

[8] Harrell J., Lee K. L., and Mark D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, 1996, *Stat Med*, 15(4): 361–87.

[9] Justice A. C., Covinsky K. E., and Berlin J. A. Assessing the generalizability of prognostic information, 1999, *Ann Intern Med*, 130(6):515–24.

[10] Copas J. B. Regression, prediction and shrinkage, 1983, *JR Stat Soc B*, 45:311–54.

[11] Efron B. and Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation, 1983, *American Statistician*, 37:36–48.

[12] Harrell J. *Regression Modelling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York, 2001.

[13] Steyerberg E. W., Bleeker S. E., Moll H. A., Grobbee D. E., and Moons K. G. Internal and external validation of predictive models: a simulation study of bias and precision in small samples, 2003, *J Clin Epidemiol*, 56(5):441–7.

[14] Steyerberg E. W., Eijkemans M. J., Harrell J., and Habbema J. D. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets, 2000, *Stat Med*, 19(8):1059–79.

[15] Steyerberg E. W., Harrell J., Borsboom G. J., Eijkemans M. J., Vergouwe Y., and Habbema J. D. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis, 2001, *J Clin Epidemiol*, 54(8):774–81.

[16] Bleeker S. E., Moll H. A., Steyerberg E. W., Donders A. R., Derksen-Lubsen G., Grobbee D. E., and Moons K. G. External validation is necessary in prediction research: a clinical example, 2003, *J Clin Epidemiol*, 56(9):826–32.

[17] Nashef S. A., Roques F., Michel P., Gauducheau E., Lemeshow S., and Salamon R. European system for cardiac operative risk evaluation (euroscore), 1999, *Eur J Cardiothorac Surg*, 16(1):9–13.

[18] Roques F., Nashef S. A., Michel P., Gauducheau E., Vincentiis C., Baudet E., Cortina J., David M., Faichney A., Gabrielle F., Gams E., Harjula A., Jones M. T., Pintor P. P., Salamon R., and Thulin L. Risk factors and outcome in european cardiac surgery: analysis of the euroscore multinational database of 19030 patients, 1999, *Eur J Cardiothorac Surg*, 15(6):816–22; discussion 822–3.

[19] Geissler H. J., Holzl P., Marohl S., Kuhn-Regnier F., Mehlhorn U., Sudkamp M., and Vivie E. R. Risk stratification in heart surgery: comparison of six score systems, 2000, *Eur J Cardiothorac Surg*, 17 (4):400–6.

[20] Gogbashian A., Sedrakyan A., and Treasure T. Euroscore: a systematic review of international performance, 2004, *Eur J Cardiothorac Surg*, 25(5):695–700.

[21] Kawachi Y., Nakashima A., Toshima Y., Arinaga K., and Kawano H. Risk stratification analysis of operative mortality in heart and thoracic aorta surgery: comparison between parsonnet and euroscore additive model, 2001, *Eur J Cardiothorac Surg*, 20(5):961–6.

[22] Michel P., Roques F., and Nashef S. A. Logistic or additive euroscore for high-risk patients?, 2003, *Eur J Cardiothorac Surg*, 23(5):684–7; discussion 687.

[23] Nashef S. A., Roques F., Hammill B. G., Peterson E. D., Michel P., Grover F. L., Wyse R. K., and Ferguson T. B. Validation of european system for cardiac operative risk evaluation (euroscore) in north american cardiac surgery, 2002, *Eur J Cardiothorac Surg*, 22(1):101–5.

[24] Nilsson J., Algotsson L., Hoglund P., Luhrs C., and Brandt J. Early mortality in coronary bypass surgery: the euroscore versus the society of thoracic surgeons risk algorithm, 2004, *Ann Thorac Surg*, 77(4):1235–9; discussion 1239–40.

[25] Yap C. H., Reid C., Yii M., Rowland M. A., Mohajeri M., Skillington P. D., Seevanayagam S., and Smith J. A. Validation of the euroscore model in australia, 2006, *Eur J Cardiothorac Surg*, 29(4): 441–6; discussion 446.

[26] Nashef S. A. Editorial comment euroscore and the japanese aorta, 2006, *Eur J Cardiothorac Surg*, 30 (4):582–3.

[27] Vergouwe Y., Steyerberg E. W., Eijkemans M. J., and Habbema J. D. Validity of prognostic models: when is a model clinically useful?, 2002, *Semin Urol Oncol*, 20(2):96–107.

[28] Peek N., Arts D. G., Bosman R. J., Voort P. H., and Keizer N. F. External validation of prognostic models for critically ill patients required substantial sample sizes, 2007, *J Clin Epidemiol*, 60(5): 491–501.

[29] Vergouwe Y., Steyerberg E. W., Eijkemans M. J., and Habbema J. D. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models, 2005, *J Clin Epidemiol*, 58(5):475–83.

[30] Oudega R., Moons K. G., and Hoes A. W. Ruling out deep venous thrombosis in primary care. a simple diagnostic algorithm including d-dimer testing, 2005, *Thromb Haemost*, 94(1):200–5.

[31] Wells P. S., Anderson D. R., Rodger M., Forgie M., Kearon C., Dreyer J., Kovacs G., Mitchell M., Lewandowski B., and Kovacs M. J. Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis, 2003, *N Engl J Med*, 349(13):1227–35.

[32] Di Nisio M., Squizzato A., Rutjes A. W., Buller H. R., Zwinderman A. H., and Bossuyt P. M. Diagnostic accuracy of d-dimer test for exclusion of venous thromboembolism: a systematic review, 2007, *J Thromb Haemost*, 5(2):296–304.

[33] Bont J., Hak E., Hoes A. W., Schipper M., Schellevis F. G., and Verheij T. J. A prediction rule for elderly primary-care patients with lower respiratory tract infections, 2007, *Eur Respir J*, 29(5):969–75.

[34] Knottnerus J. A. Between iatrotropic stimulus and interiatric referral: the domain of primary care research, 2002, *J Clin Epidemiol*, 55(12):1201–6.

[35] Wells P. S., Hirsh J., Anderson D. R., Lensing A. W., Foster G., Kearon C., Weitz J., D'Ovidio R., Cogo A., and Prandoni P. Accuracy of clinical assessment of deep-vein thrombosis, 1995, *Lancet*, 345 (8961):1326–30.

[36] Oudega R., Hoes A. W., and Moons K. G. The wells rule does not adequately rule out deep venous thrombosis in primary care patients, 2005, *Ann Intern Med*, 143(2):100–7.

[37] Stiell I. G., Greenberg G. H., McKnight R. D., Nair R. C., McDowell I., and Worthington J. R. A study to develop clinical decision rules for the use of radiography in acute ankle injuries, 1992, *Ann Emerg Med*, 21(4):384–90.

[38] Myers A., Canty K., and Nelson T. Are the ottawa ankle rules helpful in ruling out the need for x ray examination in children?, 2005, *Arch Dis Child*, 90(12):1309–11.

[39] Altman D. G. *Prognostic models: a methodological framework and review of models for breast cancer*. New York, 2007.

[40] Hosmer D. W. and S L. *Applied logistic regression*. John Wiley and Sons, Inc., New York, 1989.

[41] Ivanov J., Tu J. V., and Naylor C. D. Ready-made, recalibrated, or remodeled? issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery, 1999, *Circulation*, 99 (16):2098–104.

[42] Steyerberg E. W., Borsboom G. J., Houwelingen H. C., Eijkemans M. J., and Habbema J. D. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage, 2004, *Stat Med*, 23(16):2567–86.

[43] Janssen K. J. M., Moons K. G. M., Kalkman C. J., Grobbee D. E., and vergouwe Y. Updating methods improved the predictive performance of clinical prediction models in new patients, 2008, *J Clin Epidemiol*, 61(1):76 – 86.

[44] D'Agostino S., Grundy S., Sullivan L. M., and Wilson P. Validation of the framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation, 2001, *Jama*, 286(2): 180–7.

[45] Liu J., Hong Y., D'Agostino S., Wu Z., Wang W., Sun J., Wilson P. W., Kannel W. B., and Zhao D. Predictive value for the chinese population of the framingham chd risk assessment tool compared with the chinese multi-provincial cohort study, 2004, *Jama*, 291(21):2591–9.

[46] Stiell I., Wells G., Laupacis A., Brison R., Verbeek R., Vandemheen K., and Naylor C. D. Multicentre trial to introduce the ottawa ankle rules for use of radiography in acute ankle injuries. multicentre ankle rule study group, 1995, *Bmj*, 311(7005):594–7.

[47] Marrie T. J., Lau C. Y., Wheeler S. L., Wong C. J., Vandervoort M. K., and Feagan B. G. A controlled trial of a critical pathway for treatment of community-acquired pneumonia. capital study investigators. community-acquired pneumonia intervention trial assessing levofloxacin, 2000, *Jama*, 283(6): 749–55.

[48] Sinuff T., Adhikari N. K., Cook D. J., Schunemann H. J., Griffith L. E., Rocker G., and Walter S. D. Mortality predictions in the intensive care unit: comparing physicians with scoring systems, 2006, *Crit Care Med*, 34(3):878–85.

[49] Blackmore C. C. Clinical prediction rules in trauma imaging: who, how, and why?, 2005, *Radiology*, 235(2):371–4.

[50] Braitman L. E. and Davidoff F. Predicting clinical states in individual patients, 1996, *Ann Intern Med*, 125(5):406–12.

[51] Brehaut J. C., Stiell I. G., Visentin L., and Graham I. D. Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice, 2005, *Acad Emerg Med*, 12(10):948–56.

[52] Brehaut J. C., Stiell I. G., and Graham I. D. Will a new clinical decision rule be widely used? the case of the canadian c-spine rule, 2006, *Acad Emerg Med*, 13(4):413–20.

[53] Stiell I. G., Wells G. A., Vandemheen K. L., Clement C. M., Lesiuk H., De Maio V. J., Laupacis A., Schull M., McKnight R. D., Verbeek R., Brison R., Cass D., Dreyer J., Eisenhauer M. A., Greenberg G. H., MacPhail I., Morrison L., Reardon M., and Worthington J. The canadian c-spine rule for radiography in alert and stable trauma patients, 2001, *Jama*, 286(15):1841–8.

[54] Altman D. G. Systematic reviews of evaluations of prognostic variables, 2001, *Bmj*, 323(7306): 224–8.

[55] Riley R. D., Abrams K. R., Sutton A. J., Lambert P. C., Jones D. R., Heney D., and Burchill S. A. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future, 2003, *Br J Cancer*, 88(8):1191–8.

[56] Khan K. S., Bachmann L. M., and Riet G. Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests, 2003, *Eur J Obstet Gynecol Reprod Biol*, 108(2):121–5.

[57] Qaseem A., Snow V., Barry P., Hornbake E. R., Rodnick J. E., Tobolic T., Ireland B., Segal J., Bass E., Weiss K. B., Green L., and Owens D. K. Current diagnosis of venous thromboembolism in primary care: a clinical practice guideline from the american academy of family physicians and the american college of physicians, 2007, *Ann Fam Med*, 5(1):57–62.

[58] Qaseem A., Snow V., Barry P., Hornbake E. R., Rodnick J. E., Tobolic T., Ireland B., Segal J. B., Bass E. B., Weiss K. B., Green L., and Owens D. K. Current diagnosis of venous thromboembolism in primary care: a clinical practice guideline from the american academy of family physicians and the american college of physicians, 2007, *Ann Intern Med*, 146(6):454–8.

[59] Toll D. B., Oudega R., Bulten R. J., Hoes A. W., and Moons K. G. Excluding deep vein thrombosis safely in primary care, 2006, *J Fam Pract*, 55(7):613–8.

[60] Mulherin S. A. and Miller W. C. Spectrum bias or spectrum effect? subgroup variation in diagnostic test evaluation, 2002, *Ann Intern Med*, 137(7):598–602.

[61] Whiting P., Rutjes A. W., Reitsma J. B., Glas A. S., Bossuyt P. M., and Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review, 2004, *Ann Intern Med*, 140(3):189–202.

Summary

In many countries, the primary care physician is the first to encounter patients with symptoms and signs suggestive of DVT. They are challenged with the difficult task to discriminate patients with DVT from those without DVT. The risk of missing the diagnosis and the risk of unnecessary referral or treatment with a potential harmful therapy has to be balanced.

The American Academy of Family Physicians and the American College of Physicians recommend that prediction rules to aid in diagnosing DVT should be applied in practice. In 2005, Oudega et al. developed a diagnostic rule for safely excluding DVT in primary care. This rule discriminates 'very low' risk (non-referral) patients from patients with an increased risk of DVT. The development data showed that application of the rule may reduce the number of patient referrals substantially (23%), at the cost of an acceptable low proportion of DVT (0.7%) in non-referred patients.

Prediction rules commonly show good performance with the data from which they were developed. However, good accuracy with the development population is no guarantee the rule will be accurate for future patients. Testing the accuracy of a prediction rule in new patients is necessary before implementing it in daily patient care. In **chapter 2** we tested the accuracy of the primary care DVT rule in a new cohort of 532 primary care patients suspected of DVT. The rule items were obtained from each patient as well as the result of compression ultrasonography (CUS) serving as the reference ("gold") standard. The rule was dichotomised at the previously determined threshold of ≤ 3 (very low risk of DVT) versus ≥ 4 (increased risk of DVT). The accuracy of the rule was primarily quantified by its ability to safely exclude DVT (percentage missed DVT cases among very low risk patients). In the validation data, the rule designated 21% of all patients as very low risk of having DVT, while none of these very low risk patients had DVT. We concluded that the rule appears to be a

safe diagnostic tool for excluding DVT in patients suspected of DVT in primary care, leading to a substantial reduction of unnecessary patient referrals to secondary care and consequently of patient burden.

Promising results on such validation studies increase the confidence for applying rules in practice. However, it is well known that the relationship between disease state and the performance of a diagnostic test or rule may change according to characteristics of the patients. Physicians may thus rightly wonder whether the diagnostic accuracy of the primary care DVT rule varies. We tested the rule's efficiency (proportion of patients designated by the rule as very low risk of DVT) and safety (percentage 'missed' DVT cases among these very low risk patients) in 'high risk' groups, i.e. across 3 age groups, in men and women, and in patients with and without a history of venous thromboembolism (VTE), separately in **chapter 3**. We retrospectively analyzed data of 2086 primary care patients suspected of DVT, in whom all rule items and the result of the reference standard (compression ultrasonography) were collected. The rule's efficiency decreased with age from 38% in the relatively young ($<$ 50 years) compared to 10% in patients aged $\geq$ 70 years. The low efficiency in the elderly could be improved without compromising the safety by increasing the D-dimer threshold. The percentage of DVT among the very low risk patients was $<$ 1.5% in all subgroups. We concluded that, regardless of patient characteristics, the rule can safely exclude DVT in relevant patient subgroups in up to almost 40% of the primary care patients suspected of DVT.

In **chapter 4**, the implementation of the primary care DVT rule in clinical practice was studied in the so-called 'AMUSE study' (Amsterdam, Maastricht, Utrecht Study on the diagnosis of thrombo-Embolism). In a large prospective cohort study, consecutive primary care patients with clinically suspected DVT were managed based on the result of the rule combined with a point-of-care D-dimer test result. Patients with a score $\leq$ 3 were not referred for ultrasound and received no anticoagulant treatment; patients with a score $\geq$ 4 were referred for ultrasound and received care as usual. We evaluated the safety (the incidence of VTE in the non-referred patients during 3 months follow-up) and efficiency (proportion of patients not referred for ultrasonography) of excluding DVT using the rule. In 500 patients the score was $\leq$ 3, of whom 7 developed VTE within 3 months (1.4%; 95% CI 0.6-2.9%). In 502 patients with scores $\geq$ 4, ultrasound showed DVT in 125 (25%). Four patients with scores $\geq$ 4 and normal ultrasound results developed VTE within 3 months (1.1%, 95% CI 0.3-2.7%). This management study showed that the diagnostic rule for suspected DVT in primary care reduces the need for referral by almost 50% and is associated with a low risk of

subsequent VTE events.

This promising result encourages the use of the diagnostic rule in clinical practice. However, simple adjustments might further improve the accuracy of the diagnostic rule; i.e. reduce the proportion of missed diagnoses among non-referred patients (safety) or increase the proportion of patients who do not need to be referred for ultrasonography (efficiency). We applied three updating methods to the AMUSE cohort to determine whether adjusting the weights of the eight included predictors or adding new diagnostic predictors could further improve the accuracy of the rule **chapter 5**. The weights of the included predictors did not need to be adjusted, but inclusion of the variables 'history of DVT' and 'prolonged travelling' significantly added predictive value to the regression formula (p-value < 0.05 using the likelihood ratio test). Although these new diagnostic variables were independent predictors of the presence of DVT, adding these to the rule did not confer the clinically relevant effects (safety and efficiency). In fact, at equal safety (1.4% missed diagnoses among the non-referred patients), the efficiency was lower (43.5%) when using the updated rule compared to the original rule (49.4%). We concluded that the diagnostic rule for excluding DVT in primary care has optimal accuracy in its original form and does not need any adjustments, providing the final evidence that the original strategy can be used to safely exclude clinically suspected DVT in primary care.

In **chapter 6** we used the data of the AMUSE cohort to estimate whether the diagnostic rule - which was deemed to be safe for excluding DVT in primary care - was also cost-effective as compared to two 'care as usual' (i.e. hospital based strategies). A Markov model with a five year time horizon was used to compare the use of the primary care DVT rule to hospital based strategies. Probabilities were derived from data collected in the AMUSE study and the literature. Societal costs and health state utilities were used, one way and probabilistic sensitivity analyses were conducted, and cost-effectiveness acceptability curves were constructed. The primary care DVT rule had both slightly lower costs and less QALYs than the care as usual strategies. We compared the primary care DVT rule to the best hospital strategy, which resulted in a saving of €138, and a QALY loss of 0,002. The iCER is €56,436. The cost-effectiveness acceptability curves showed that the primary care DVT rule has the highest probability of being cost-effective, even exceeding ceiling ratios over €80,000. We concluded that the primary care DVT rule is not only safe, but also cost-effective as compared to hospital based strategies.

The use of a D-dimer assay is a little troublesome, because there is no standardis-

ation in the test results and cut-off values for the different assays. This limits the comparison of the diagnostic accuracy of various D-dimer assays. Both, manufacturers and investigators, often consider the D-dimer test results normal if the detected concentration is $<$ 500 ng/ml. Little evidence however is available if this threshold is indeed the optimal threshold to exclude DVT in primary care, and whether different D-dimer assays require different thresholds. In **chapter 7**, we evaluated two sensitive D-dimer assays, a rapid ELISA method (VIDAS, bioMérieux, France) and a quantitative latex assay method (Tinaquant, Roche, Germany). Both manufacturers address 500 ng/ml as cut-off value for (ab)normality but carefully added that each laboratory should establish its own normal range based upon the population tested. We investigated for both assays if this threshold for excluding DVT is of value in clinical practice of primary care and whether other thresholds might be preferred. Given the main purpose of D-dimer testing in primary care - i.e. to exclude DVT - we confirm the cut-off value of 500 ng/ml for the VIDAS assay. In contrast, for the Tinaquant assay our data suggest to use a lower threshold than 500 ng/ml. The 500 ng/ml threshold yields an unacceptably low sensitivity and negative predictive value. Cut-off values for the Tinaquant assay in the range of 200 to 400 ng/ml, depending on whether the assay is used as stand-alone test or in combination with clinical characteristics, yielded better diagnostic accuracy. But also at these cut-off values, the diagnostic accuracy of the Tinaquant assay, remained inferior to the VIDAS assay when applied in primary care.

To enable the exclusion of DVT, by using a diagnostic rule and D-dimer testing, in the near-patient situation -without the need to refer a patient for D-dimer testing- several so-called 'point of care' (POC) D-dimer assays have recently been introduced. POC D-dimer assays can directly be performed on whole blood with the use of hardly any additional laboratory handling. However, POC D-dimer tests can only be safely substituted for conventional D-dimer assays, if the POC assays have acceptable test characteristics. We determined the diagnostic accuracy and user-friendliness of five different POC D-dimer assays in a cohort of 200 primary care patients suspected of DVT (**chapter 8**). At the manufacturers' provided thresholds, all D-dimer assays showed negative predictive values higher than 98%. Vidas® assay had a sensitivity of 100%, with a specificity of 40%. Pathfast^{TM}, Triage® and Cardiac D-dimer assays showed a equally good sensitivity of 95.8%, while specificities were 35%, 48%, and 57%, respectively. The Simplify® D-dimer assay had a sensitivity of 91.7% with a specificity of 63%. Regarding the user-friendliness of the assays, differences were mainly determined by (the need for) calibration, time needed to obtain a D-dimer assay result, portability of the device, and the retail prizes. In conclusion, all POC D-

Dimer assays showed reasonable to good diagnostic accuracy to aid in the diagnosis of DVT. The user-friendliness of the Vidas® and Pathfast^{TM} assays may be limited for use in primary care, in contrast to that of the Simplify®, Cardiac and Triage® assays.

The general discussion consists of two parts. **Chapter 9.1** is a review describing important aspects of validation studies and updating methods, impact analyses and the implementation of prediction rules. The consecutive phases in multivariable prediction research as described in chapter 9.1, is briefly illustrated with the diagnostic rule that has been described in chapters 1 to 6 (**chapter 9.2**).

In validation studies, temporal, geographical, and domain validation can be distinguished. Temporal validation tests the generalizability of a prediction rule 'over time'. Geographical validation studies typically test the generalizability of a prediction rule in a patient population that is similarly defined as the development population, though in other geographical areas. Domain validation is the broadest form of validation, and tests the generalizability of a prediction rule across different domains, such as patients from a different setting (primary, secondary or tertiary), inpatients versus outpatients, patients of different age categories (e.g. adults versus adolescents or children), of a different gender, and perhaps from a different type of hospital (academic versus general hospital). In general, the potential for differences between the derivation and validation population is smallest in a temporal validation study, and largest in a domain validation study. Consequently, good results on a domain validation study are considered to provide the strongest evidence that the prediction rule can be generalized to new patients. Understandably, prediction rules that fail to maintain their accuracy in new patients should not be applied in clinical practice. However, these rules should not directly be rejected, because adjusting the prediction rules (updating) has many advantages compared to developing new rules. Various updating methods that can be applied to improve the predictive performance are discussed. The need for impact analyses to assess the true effect of the implementation of a prediction rule in daily clinical practice is discussed, and the barriers during the implementation of a prediction rule in clinical practice are considered. The review concludes with remaining methodological issues in prediction research.

Samenvatting

Jaarlijks krijgen 32.000 mensen in Nederland een diep veneuze trombose (DVT), ook wel trombosebeen genoemd. DVT ontstaat als er een stolsel wordt gevormd in de grote aderen van de benen. Stukjes van het stolsel kunnen afbreken en via de bloedbaan naar de longen worden vervoerd. Hier kunnen de stolsels een longembolie veroorzaken, wat vervolgens kan leiden tot de dood. Anderzijds brengt het (onnodig) behandelen van DVT risico's met zich mee: de bloedverdunners kunnen ernstige bloedingen veroorzaken.

Een rood, gezwollen en/of pijnlijk onderbeen kan duiden op een DVT, maar ook op vele andere (minder ernstige en minder acute) aandoeningen. De huisarts moet beslissen welke van deze patiënten doorverwezen worden naar het ziekenhuis voor belastend en duur onderzoek. Tot voorkort was deze beslissing niet eenvoudig. De huisarts stuurde uit veiligheidsoverwegingen (bijna) alle patiënten die mogelijk DVT hebben door naar het ziekenhuis, terwijl meer dan driekwart van de patiënten uiteindelijk geen DVT blijkt te hebben. Dit is vervelend voor de patiënt en kost de maatschappij onnodig veel geld. Om huisartsen een handvat te bieden bij de diagnostiek van DVT, hebben Oudega en collega's een eenvoudige diagnostische beslisregel ontwikkeld. De regel bestaat uit 7 predictoren uit anamnese en lichamelijk onderzoek plus een D-dimeer test uitslag, en onderscheidt patiënten met een laag risico op DVT (geen verwijzing voor echografie) van patiënten met een verhoogd risico op DVT (verwijzen voor echografie). In het onderzoek van Oudega zou toepassing van de regel het aantal verwijzingen met 23% verminderen (proportie laag risico patiënten), terwijl slechts 0.7% van de laag risico patiënten DVT had.

Diagnostische regels zijn meestal accuraat in de data waarin ze zijn ontwikkeld (ontwikkelingsdata). Goede accuratesse in de ontwikkelingsdata is echter geen garantie voor goede voorspellingen in nieuwe patiënten. De accuratesse van nieuw ontwikkelde

regels moet dan ook eerst getest worden in nieuwe patiënten, voordat ze worden toegepast in de klinische praktijk. In **hoofdstuk 2** hebben we de accuratesse van de diagnostische beslisregel getest in een nieuw cohort van 532 opeenvolgende eerstelijns patiënten met een verdenking op DVT. Nadat de predictoren van de regel verzameld waren, werden alle patiënten verwezen voor echo-onderzoek (referentie standaard). De score (range 0-14) werd gedichotomiseerd op de afkapwaarde die in eerder onderzoek was bepaald: ≤ 3 (laag risico op DVT) versus ≥ 4 (verhoogd risico op DVT). We bepaalden de veiligheid van de regel (proportie gemiste DVTs onder de laag risico patiënten) als primaire maat van accuratesse. De regel categoriseerde 21% van alle patiënten als laag risico op DVT. Geen van de laag risico patiënten had DVT. We concludeerden dat de regel een veilig diagnostisch handvat voor het uitsluiten van DVT in de huisartsenpraktijk lijkt, en het aantal onnodige verwijzingen naar het ziekenhuis substantieel kan reduceren.

De goede accuratesse van de diagnostische beslisregel in nieuwe patiënten vergroot het vertrouwen om de regel in de praktijk toe te passen. Echter, het is bekend dat de associatie tussen de ziekte status en de accuratesse van een diagnostische regel kan variëren afhankelijk van patiëntkarakteristieken. Artsen zouden daardoor mogelijk terughoudend kunnen zijn in het gebruik van de diagnostische beslisregel in bepaalde patiëntgroepen. In **hoofdstuk 3** werd de efficiëntie (proportie patiënten met een laag risico volgens de regel) en de veiligheid (% gemiste DVTs onder de laag risico patiënten) van de regel bepaald in 3 leeftijdsgroepen, in mannen en vrouwen apart, en in patiënten met en zonder een verleden van veneuze tromboemolie (VTE). Retrospectief werd data van 2086 eerstelijns patiënten met verdenking op DVT, van wie alle predictoren van de regel en de referentie standaard (echografie) bekend waren, geanalyseerd. De efficiëntie van de regel daalde met toenemende leeftijd van 38% in relatief jonge patiënten ($<$50 jaar), ten opzichte van 10% in patiënten $\geq$ 70 jaar. De lage efficiëntie van de regel in ouderen kan verbeterd worden -zonder de veiligheid van de regel in het geding te brengen- door de D-dimeer afkapwaarde te verhogen. Het percentage DVT onder de laag risico patiënten was $<$1.5% in alle subgroepen. We concludeerden dat de diagnostische beslisregel DVT veilig kan uitsluiten, ongeacht patiëntkarakteristieken.

In **hoofdstuk 4** hebben we de implementatie van de diagnostische beslisregel bestudeerd in het zogenaamde 'AMUSE' onderzoek (Amsterdam, Maastricht, Utrecht Study on the diagnosis of thromboEmbolism). In dat onderzoek gebruikten huisartsen de regel (inclusief een 'point-of-care' D-dimeer test) om bij een groot aantal opeenvolgende patiënten met verdenking DVT te bepalen of verwijzing voor echografie

geïndiceerd was. Patiënten met een score ≤ 3 werden niet verwezen voor echografie en werden niet behandeld met anticoagulantia; patiënten met een score ≥ 4 werden verwezen voor echografie en kregen gebruikelijke zorg. De veiligheid (incidentie VTE onder de niet verwezen patiënten binnen de 3 maanden follow-up periode) en efficiëntie (proportie patiënten niet verwezen voor echografie) van de diagnostische regel werd bepaald. Van de 500 patiënten met een score ≤ 3, ontwikkelden 7 VTE binnen de 3 maanden follow-up (1.4%; 95% CI 0.6-2.9%). Van de 502 patiënten met een score ≥ 4 was de echo positief in 125 (25%). Drie patiënten met een negatieve echo ontwikkelden VTE binnen de 3 maanden follow-up (1.1%, 95% CI 0.3-2.7%). We concludeerden dat de diagnostische regel voor het uitsluiten van DVT in de huisartsenpraktijk het aantal verwijzingen voor echografie met bijna 50% kan reduceren, ten kosten van slechts een laag risico VTE onder de niet verwezen patiënten.

Het resultaat van de AMUSE onderzoek zou het gebruik van de diagnostische regel in de praktijk kunnen stimuleren. Echter, eenvoudige aanpassingen zouden de accuratesse van de regel verder kunnen verbeteren; d.w.z. de proportie gemiste diagnoses onder niet-verwezen patiënten verlagen (veiligheid) of de proportie patiënten die niet verwezen hoeven te worden vergroten (efficiëntie). In **hoofdstuk 5** hebben we - met behulp van 'state of the art' methodologie voor het updaten van klinische predictie regels - onderzocht of de accuratesse van de diagnostische regel inderdaad nog verder verbeterd kan worden op basis van de data van het AMUSE onderzoek. De update methodes werden toegepast op de originele, onderliggende regressie formule van de (gesimplificeerde) diagnostische regel. De regressiecoëfficienten van de individuele predictoren van de diagnostische regel behoefden geen aanpassing. Uitbreiding van de regel met de variabelen 'DVT in het verleden' en 'langdurig reizen' verhoogden de voorspellende waarde van de regressie formule significant ($p < 0.05$, likelihood ratio test). Ondanks dat deze diagnostische variabelen onafhankelijke predictoren van de aanwezigheid van DVT zijn, verbeterde de klinisch relevante maten niet na toevoegen van de variabelen: nadat de ge-update regressie formule was getransformeerd in een ge-update diagnostische regel, waren de veiligheid en de efficiëntie van de diagnostische regel niet verbeterd. Bij een gelijke veiligheid (1.4% gemiste diagnoses onder niet-verwezen patiënten), was de efficiëntie van de ge-update regel (43.5%) zelfs slechter dan de originele diagnostische regel (49.4%). We concludeerden dat de originele diagnostische regel optimale accuratesse heeft voor het veilig uitsluiten van DVT in de huisartsenpraktijk en geen aanpassingen behoeft.

In **hoofdstuk 6** hebben we de data van het AMUSE cohort gebruikt om te bepalen of de diagnostische regel - welke veilig bleek voor gebruik in de huisartsenpraktijk

- ook kosten-effectief was in vergelijking met twee 'gebruikelijke zorg' (ziekenhuis) strategieën. Een Markov model met een vijf jaar tijdshorizon werd gebruikt om de regel te vergelijken met de ziekenhuis strategieën. De data van het AMUSE cohort en de literatuur werden gebruikt om modelparameters te definiëren. De regel had ietwat lagere kosten en minder QALYs dan de ziekenhuis strategieën. Het gebruik van de regel resulteerde - in vergelijking met de beste ziekenhuis strategie - in een besparing van €138 en een QALY verlies van 0,002. De iCER was €56,436. We concludeerden dat het gebruik van de diagnostische regel niet alleen veilig was, maar ook kosten-effectief in vergelijking tot ziekenhuis strategieën.

Het gebruik van een D-dimeer test is niet eenduidig omdat er geen standaardisatie is van het testresultaat en evenmin van de drempelwaarde van de verschillende testen. Dit beperkt het vergelijken van de diagnostische accuratesse van verschillende D-dimeer testen. Zowel fabrikanten als onderzoekers beschouwen een D-dimeer test resultaat als normaal als de geteste waarde kleiner is dan 500 ng/ml. Er is echter weinig bewijs beschikbaar of dit inderdaad de optimale drempel is om DVT uit te sluiten in de huisartsenpraktijk. Het is evenmin duidelijk of verschillende testen om verschillende drempelwaarden vragen. In **hoofdstuk 7** onderzochten we twee sensitieve D-dimeer testen, een snelle ELISA test (VIDAS) en een kwantitatieve latex test (Tinaquant). De fabrikanten van beide testen geven 500 ng/ml aan als afkapwaarde voor een (ab)normale testuitslag, maar geven tevens zorgvuldig aan dat ieder laboratorium zelf de normaalwaarden dient te bepalen bij de patiënten waarvoor de test gebruikt wordt. We hebben van beide testen onderzocht of deze drempelwaarde om DVT uit te sluiten gebruikt zou moeten worden door de huisarts of dat andere drempelwaarden wellicht de voorkeur hebben. Voor de VIDAS test bleek de afkapwaarde van 500 ng/ml inderdaad optimaal om DVT uit te sluiten. Daarentegen suggereerde onze gegevens dat voor de Tinaquant een lagere drempelwaarde dan 500 ng/ml geschikter is, aangezien de sensitiviteit bij een drempel van 500 ng/ml onacceptabel laag is (91%). Echter, ook indien de drempelwaarde voor de Tinaquant werd verlaagd, gaf de VIDAS betere diagnostische accuratesse.

Verschillende zogenaamde 'point-of-care' (POC) D-dimeer testen zijn recentelijk ontwikkeld, die direct en (relatief) eenvoudig uitgevoerd kunnen worden op veneus of capillair bloed. Een D-dimeer test resultaat kan daardoor verkregen worden zonder de patiënt te verwijzen voor laboratorium onderzoek. Het gebruik van POC D-dimeer testen zou de praktische toepasbaarheid van de diagnostische regel in de huisartsenpraktijk sterk verbeteren. Echter, de conventionele D-dimeer testen kunnen alleen vervangen worden door POC D-dimeer testen, als deze goede test karateristieken

hebben. In **hoofdstuk 8** hebben we de diagnostische accuratesse en gebruiksvriendelijkheid van vijf verschillende POC D-dimeer testen in huisartspatiënten met verdenking DVT bepaald. Wanneer de door de fabrikant afgegeven drempelwaardes werden gebruikt, hadden alle D-dimeer testen negatief voorspellende waardes hoger dan 98%. De Vidas® test had een sensitiviteit van 100% en een specificiteit van 40%. Pathfast^{TM}, Triage® en Cardiac D-dimeer testen hadden allemaal een sensitiviteit van 95.8%, terwijl de specificiteit respectievelijk 35%, 48% en 57% was. De Simplify® D-dimeer test had een relatief lage sensitiviteit (91.7%) en een hoge specificiteit (63%). De gebruiksvriendelijkheid van de verschillende testen verschilde voornamelijk in calibratie, tijdsduur voor het bepalen van één D-dimeer test resultaat, draagbaarheid van het apparaat, en de verkoopprijzen. We concludeerden dat alle POC D-dimeer testen (redelijk) goede accuratesse hebben voor het uitsluiten van DVT. De gebruiksvriendelijkheid van de Vidas® en Pathfast^{TM} zou beperkt kunnen zijn voor gebruik in de huisartsenpraktijk, in tegenstelling tot die van de Simplify®, Cardiac en Triage® D-dimeer testen.

De algemene discussie bestaat uit twee delen. In **hoofdstuk 9.1** geven we een overzicht van belangrijke aspecten van validatie studies en updatemethoden, impact analyses, en de implementatie van predictie modellen in de praktijk. In het tweede deel (**hoofdstuk 9.2**), worden de opeenvolgende fasen van multivariabele predictie onderzoek zoals beschreven in hoofdstuk 9.1 kort toegelicht aan de hand van de diagnostische regel die beschreven is in hoofdstukken 1 tot en met 6.

Temporele, geografische en domein validaties kunnen worden onderscheiden. Een temporele validatie test de generaliseerbaarheid van een model in een andere periode. Een geografische validatie test de generaliseerbaarheid van een model in andere ziekenhuizen, instituten of in andere geografische gebieden. Een domein validatie test de generaliseerbaarheid in bijvoorbeeld patiënten uit een andere setting (eerste, tweede of derde lijn), patiënten uit andere leeftijdscategorieën (bijvoorbeeld volwassenen versus tieners of kinderen), patiënten van een andere sekse of patiënten uit een ander type ziekenhuis (academisch versus perifeer). Meestal zijn de verschillen tussen de ontwikkelingspopulatie en de validatie populatie het kleinst bij een temporele validatie, en het grootst bij een domein validatie. Daarom is een goede domein validatie een beter bewijs dat een model generaliseerbaar is naar nieuwe patiënten dan een temporele validatie. Modellen met beperkte generaliseerbaarheid moeten uiteraard niet worden gebruikt in de klinische praktijk. Echter, deze modellen zouden ook niet direct verworpen moeten worden, aangezien het aanpassen aan deze modellen (updaten) vele voordelen heeft boven het ontwikkelen van nieuwe modellen. Verschillende update methodes voor het verbeteren het voorspellende vermogen

van een model worden beschreven. Ook de noodzaak tot het uitvoeren van impact analyses om het ware effect van een model te meten wordt besproken, en de barrières die men tegenkomt tijdens het implementeren van een model. Het hoofdstuk wordt afgesloten met enkele toekomstige methodologische uitdagingen in predictie onderzoek.

Dankwoord

Het is een cliché, maar een promotieonderzoek doe je niet alleen. Er zijn veel mensen die in meer of mindere mate belangrijk zijn geweest voor het tot stand komen van dit proefschrift, die ik hier graag wil bedanken.

Allereerst wil ik graag alle huisartsen en assistentes die aan het AMUSE onderzoek hebben meegewerkt danken voor hun inzet. In de afgelopen drie jaar is het me duidelijk geworden dat onderzoek in de huisartsenpraktijk vaak 'even niet zo goed uitkomt', maar desondanks hebben jullie toch meer dan 1000 patiënten geïncludeerd!

De laboratoria van Rotterdam (STAR-MDC), Deventer en Groningen (LabNoord) hartelijk dank voor de inspanningen omtrent de AMUSE-POC studie.

En natuurlijk dank aan alle patiënten die bereid waren aan de onderzoeken deel te nemen.

Behalve goede data is goede begeleiding natuurlijk onontbeerlijk om je promotietijd fatsoenlijk door te komen! Beste (co-)promotoren, ik heb me de afgelopen 3 jaren gelukkig geprezen met jullie driemanschap aan begeleiding!

Carl, het charmeert je dat je graag laagdrempelig bent ("loop anders gewoon even binnen"). Maar sinds ik aangesteld ben als promovenda, ben jij benoemd tot prof. en is je agenda geëxplodeerd! Ook al was het praktisch vaak niet mogelijk om even bij je binnen te lopen, was ik meer dan tevreden met je enthousiaste en positieve manier van begeleiden. Het was erg fijn dat je -ondanks je druk bezette agenda- bijna altijd snel tijd kon vrijmaken om mijn stukken nauwkeurig te becommentariëren en te bespreken.

Arno, ik weet me nog goed te herinneren dat jij tijdens mijn sollicitatiegesprek vol overtuiging zei dat onderzoek doen helemaal niet leuk is! Ik sprak dat toen nog tegen, maar ik moet toegeven dat er wel een paar momenten zijn geweest tijdens mijn

promotie (een jaar tevergeefs gewerkt aan een diagnostische meta-analyse omdat een andere onderzoeker hetzelfde onderzoek net iets eerder publiceerde of frustraties over een 'imperfecte' database) dat ik je gelijk moest geven. Gelukkig hadden die momenten zeker niet de overhand, en daar heeft de prettige begeleiding van jouw kant ook zeker aan bijgedragen!

Ruud, na 25 jaar als huisarts te hebben gewerkt heb je je helemaal overgegeven aan de wetenschap! Het netwerk dat je hebt opgebouwd tijdens je jaren als huisarts bleek zeer waardevol voor het AMUSE onderzoek. Jouw inzet voor het logistieke reilen en zeilen van het onderzoek en je enthousiasme om het gebruik van "jouw regel" verder wetenschappelijk te onderbouwen waren onmisbaar voor het succesvol afronden van het AMUSE onderzoek.

Constance, Lara, Annina en Monique dank jullie wel voor het plannen, afzeggen en opnieuw plannen van de afspraken met Carl en Arno. Lara, jij hebt tijdens je korte tijd als secretaresse bij Carl ook nog veel werk voor me verzet door data in te voeren, hartelijk dank daarvoor!

Zoals de naam AMUSE ('Amsterdam Maastricht Utrecht Study on the diagnosis of thromboEmboly') al suggereert was "mijn" onderzoek een samenwerkingsproject. Amsterdam (AMC) was vertegenwoordigd door Harry, Henk en Eit Frits, en Maastricht was vier 'man' sterk: Martin, Jelle, Manuela en Arina. Bedankt dat jullie elke twee maanden weer naar Utrecht afreisden voor overleg. Het was prettig met jullie samenwerken en ik vond het heel bijzonder om met zo'n grote groep 'het artikel' in één dag te schrijven! Arina en Eit Frits: succes met het afronden van jullie proefschrift.

Yvonne en Kristel, jullie zijn beide co-auteur op 2 van de publicaties in dit proefschrift. Bedankt voor jullie bijdrage en de fijne samenwerking.

Geert-Jan, als opvolger op het AMUSE-2 project heb je ook een belangrijke bijdrage geleverd aan het afronden van "mijn" onderzoek. Je bent een gezellige collega. Succes met het verder uitpluizen van de AMUSE-1 data en het verzamelen van je "eigen" data.

Ruben, René en Marloes, jullie hebben als stagiaires op het AMUSE project alledrie veel en goed werk verzet! Jullie bijdrage blijkt onder andere uit een co-auteurschap in een van de publicaties in dit boekje. Bedankt voor de inzet en succes met jullie carrière!

Mariëtte, fijn dat je me een aantal maanden hebt ondersteund bij de administratieve klusjes rondom de data-verzameling. Ik hoop dat je je draai nu helemaal hebt gevonden in Leiden.

Predictieclub: Carl, Yvonne, Kristel, Corné, Jolanda, Martijn, Eva, Lidewij, Peter, Michael, Teus, Geert-Jan, Joris, Mireille, Sjoerd en Linda (het worden er haast te veel om op te noemen) bedankt voor de leerzame en gezellige predictieclub-uurtjes. En natuurlijk voor de gezellige etentjes (worden oud-leden ook nog eens uitgenodigd?)!

Alle collega-promovendi: bedankt voor de gezelligheid tijdens de lunches, kameruitjes, borrels, (promovendi)-uitjes, squash-dates en gewoon tussen het werken door. Mijn kamergenootjes van het eerste uur van kamer 6.104 wil ik graag in het bijzonder bedanken. Maud, Esther, Anne, Wijnand, Yvonne, Manja en Sjors jullie hebben er voor gezorgd dat ik me meteen 'thuis' voelde op het Julius. De gezellige sfeer werd -zowel tijdens als na het werk- goed voortgezet met de komst van Carline, Soner, Ingrid, Madeleine en Geert-Jan. Maud en Esther super dat we de laatste lootjes van onze promotie samen op kamer 6.103 hebben afgelegd! Jullie mogen nog even doorbikkelen.... Ik hoop jullie na mijn promotie nog te blijven zien.

Beste Dennis, zoals je zelf zegt "af en toe moet je iemand gelukkig maken" en dat is je zeker gelukt met het ontwerp van de kaft. Bedankt!

Iedereen die heeft gezorgd voor de broodnodige ontspanning na de onderzoekuurtjes wil ik hier ook graag bedanken. Aniek, Femke en Sanne, de jaarlijkse weekendjes en de gezellige avondjes houden we erin!

Oud-fyrfadders, het contact is (natuurlijk) niet meer zo frequent als in Maastricht, maar de gezelligheid blijft! Minimaal elk jaar Hajraa natuurlijk! Aukje, gezellig dat je me vanuit Maastricht -zij het met een kleine omweg- gevolgd hebt naar Utrecht. Onze sport-dates willen nog wel eens geannuleerd worden, maar de gezellige etentjes gaan gelukkig wel altijd door! Binnenkort maar weer het jaarlijkse weekendje zeilen vastleggen?

Tom, we zitten nu min of meer in hetzelfde schuitje. Jij aan het begin, ik aan het eind! Wie weet kan ik je nog eens wijs advies geven :-). En anders houden we het gewoon bij een potje kolonisten of carcassonne (het wordt weer eens tijd!).

Beste Paranimfen, Maud en Aniek, ook al heeft het enkel een ceremoniële betekenis, ik vind het fijn dat jullie achter mijn zullen staan! Maud, met veel plezier heb

ik vanaf dag één een 'bureau-eiland' met je mogen delen. Na dik twee jaar op 6.104 verhuisden we samen naar 6.103 en hebben daar de boel verbouwd zodat we weer samen aan een eiland konden zitten. Die bureaus tegen de muur en met ruggen naar elkaar leek ons toch echt veel te ongezellig. Ik zal mijn bureau-maatje gaan missen!

Aniek, we zijn al bevriend sinds de peuterspeelzaal, en hebben wat dat betreft een PR op ons naam. Ik hoop dat we nog vele jaren aan dat record gaan toevoegen.

(Schoon)familie, bedankt voor jullie interesse. Ik beloof, vanaf nu geen 'smoesjes' meer van mij om geen surprise te maken met sinterklaas of om geen gerecht voor te bereiden met kerst.

Lieve Ralf, jij kent de perikelen van promoveren als geen ander. In jouw dankwoord schreef je nog dat je hoopt een positieve bijdrage te leveren aan mijn promotie. Nou, dat heb je zeker gedaan, en niet alleen door de inhoud zo mooi te lay-outen! Super bedankt!

Curriculum Vitae

Diane Toll was born on May 15^{th}, 1981, in 's-Hertogenbosch, the Netherlands. After graduating secondary school in 1999 at Gymnasium Beekvliet in Sint-Michielsgestel, she obtained her propaedeutics Business Economics at the Hogeschool 's-Hertogenbosch. In 2000, she started her training in Health Sciences at Maastricht University. During this period, she conducted a research project at the department of Human Biology, faculty of Health Sciences at Maastricht University, under supervision of Prof.dr. K.R. Westerterp and G. Plasqui. For her master thesis, entitled "Estimating Energy Expenditure from Accelerometer Output", she received the Master Thesis Award 2004 of the Faculty of Health Sciences. In august 2004, she obtained her Master of Science degree in Health Sciences, specialization Biological Health Sciences (with merit). Subsequently, she worked as a junior researcher at the department of Human Biology, faculty of Health Sciences at Maastricht University. In February 2005, she started working as a PhD-student on the studies described in this thesis at the Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht (supervised by Prof.dr. K.G.M. Moons, Prof.dr. A.W. Hoes, and dr. R. Oudega). She obtained her Master of Science in Epidemiology at Utrecht University in August 2007 (cum laude). As of February 2008, she is working as a junior research fellow at TNO Quality of Life, Leiden.