

Validity and diagnostic value of tests used in the diagnostic work-up of haemostatic disorders

Citation for published version (APA):

Nagler, M. (2014). *Validity and diagnostic value of tests used in the diagnostic work-up of haemostatic disorders*. [Doctoral Thesis, Maastricht University]. Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20141203mn>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/dis.20141203mn](https://doi.org/10.26481/dis.20141203mn)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Validity and diagnostic value of tests
used in the diagnostic work-up
of haemostatic disorders

ISBN 978 94 6159 389 4
Datawyse / Universitaire Pers Maastricht

© copyright M Nagler, Maastricht 2014

Validity and diagnostic value of tests used in the diagnostic work-up of haemostatic disorders

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Professor dr. L.L.G. Soete
volgens het besluit van het College van Decanen, in het openbaar te verdedigen op
woensdag 3 december 2014 om 12.00 uur

door

Michael Nagler

geboren te Nordhausen, Duitsland, op 22 maart 1974



Supervisor

Prof. dr. H. ten Cate

Co-supervisor

Prof. dr. W.A. Wullemin (University of Berne)

Prof. dr. L.M. Bachmann (University of Zurich)

Assessment Committee

Prof. dr. M.H. Prins (voorzitter)

Prof. dr. A. Angelillo-Scherrer (University of Berne)

Prof. dr. J.C.M. Meijers (University of Amsterdam)

Prof. dr. A.E. Marcus

Dr. H.M.H. Spronk

TABLE OF CONTENTS

| | | |
|------------------|---|-----|
| Chapter 1 | General Introduction | 7 |
| Chapter 2 | Technical efficacy of thromboelastometry and thrombelastography analysis in the evaluation of haemostasis: A Systematic Review | 15 |
| Chapter 3 | Consistency of thromboelastometry analysis under scrutiny: Results of a systematic evaluation within and between analysers | 27 |
| Chapter 4 | Impact of changes in haematocrit level and platelet count on thromboelastometry parameters | 41 |
| | Appendix – Thromboelastometry changes in myeloproliferative neoplasms – surrogate for a procoagulant haemostatic imbalance or a consequence of technical reasons? | 53 |
| Chapter 5 | Accuracy of the point-of-care coagulometer CoaguChek XS in the hands of patients | 57 |
| Chapter 6 | Prospective evaluation of the interobserver reliability of the 4T's score in patients with suspected heparin-induced thrombocytopenia | 63 |
| | Appendix – Diagnostic value of the 4T's score for heparin-induced thrombocytopenia in the critically ill | 73 |
| Chapter 7 | Variability between laboratories performing coagulation tests with identical platforms: a nationwide evaluation study | 77 |
| Chapter 8 | Methodological shortcomings and design-related bias in studies investigating diagnostic tests for venous thromboembolic diseases – Literature survey | 89 |
| Chapter 9 | General Discussion | 115 |
| | Summary | 123 |
| | Samenvatting | 127 |
| | Valorization addendum | 131 |
| | Dankwoord | 135 |
| | Curriculum vitae | 137 |
| | Publications | 138 |

CHAPTER 1

General Introduction

The purpose of haemostasis

Haemostasis protects us from harm of bleeding and thrombosis. An intact haemostasis generates a blood clot at the site of an injured vessel, and simultaneously prevents us against any non-physiologic thrombosis. Efficient haemostasis requires an orchestrated and extremely fascinating interplay of blood vessels, platelet plug formation, clotting cascade, and fibrinolysis [1]. Over the last years, I have been a student of haemostasis and corresponding laboratory tests and the thesis at hand summarises part of my work in this context.

Following some introductory remarks on hereditary and acquired disorders of haemostasis, this introduction provides a short overview of laboratory testing in clinical practice, address issues of test validity and presents an overview of my thesis.

Disorders of haemostasis contribute to the world's health burden

Any hereditary or acquired disorder of haemostasis may lead to bleeding or venous and arterial thromboembolism, which represents a significant health burden worldwide [2,3].

Bleeding is the most common reason for trauma deaths [4] and thus a common cause of death in young adults in developed countries [5]. On the other hand, anticoagulants, such as warfarin, bear an inherent risk for mortality in elderly trauma patients [6] and are regarded as the most common reason for acquired disorders of haemostasis. Moreover, commonly prescribed cardiovascular drugs lead to a disordered platelet function increasing the propensity of bleeding [7,8]. Finally, although hereditary bleeding disorders such as haemophilia or von Willebrand disease are much less common, the expenditures for treating these patients are substantial [9].

Haemostatic changes are considered as prerequisite for development of venous and arterial thromboembolism [10]. Furthermore, altered coagulation factors are regarded as relevant risk factor for venous thromboembolism [11,12] and possibly for cardio-vascular disease [13].

Laboratory tests used in diagnostic workup and monitoring therapy

Laboratory tests are of paramount importance in the management of any thromboembolic and bleeding disorder. For instance, the D-dimer levels are a cornerstone in the exclusion of venous thromboembolism [14,15] and also an important prognostic parameter for recurrent thrombosis [16,17]. Prothrombin time, activated partial thromboplastin time, thrombin time, ecarin clotting time and anti Xa assays are prerequisite

for monitoring therapy of anticoagulants [18,19]. In recurrent thrombosis, a number of laboratory parameters such as antiphospholipid antibodies, factor V Leiden mutation, prothrombin gene mutation, protein C and protein S level, fibrinogen, and antithrombin [20] assist clinicians assessing the risk in the individual patient. Platelet function tests and determination of the level of the individual coagulation factors are needed to evaluate patients with a bleeding tendency [21].

In recent years, global coagulation parameters and point-of-care devices such as thromboelastometry and thrombelastography for rapid determination of a disturbed haemostasis in patients with massive bleeding have entered the market [22]. Determination of PF4/heparin antibodies is essential for work-up in patients with the critical situation of suspected heparin-induced thrombocytopenia [23]. Platelet function tests are increasingly used for monitoring drugs used to inhibit platelet function in cardiovascular and cerebrovascular disease [24]. Combinations of several tests are required for work-up of complex situations, such as disseminated intravascular coagulation and liver dysfunction [25].

Validity and clinical value of laboratory tests

Valid, accurate and precise laboratory tests for diagnostic workup and monitoring therapy may help preventing thromboembolic complications and bleeding. Missing or delaying a diagnosis of deep vein thrombosis, pulmonary embolism or heparin-induced thrombocytopenia may lead to serious complications and death [15,23,26]. A precise intensity of anticoagulation, as guided by the respective laboratory tests, improves efficacy and reduces bleedings [27-29]. Furthermore, rapid detection of coagulopathy in trauma patients prevents (unnecessary) transfusions [30]. However, a number of authors have identified serious gaps in knowledge of validity, accuracy and precision as well as clinical value for a number of laboratory methods, in various situations [17,30-37]. These limitations result in uncertainty with regard to diagnosis, prognosis and monitoring therapy, probably leading to application of unnecessary treatments, withhold of effective treatments or non-precise monitoring of drugs. Thus, a better understanding of validity and clinical value of laboratory tests used in haemostasis not only may help to improve patient care but also to limit health costs.

Aim and outline of the thesis

In this thesis, we aimed to investigate the validity and accuracy of haemostasis laboratory tests applied in a broad range of conditions. In **Chapter 2**, the technical efficacy of thromboelastometry (ROTEM) and thrombelastography (TEG) analysis against estab-

lished parameters of haemostasis is discussed. A systematic review was conducted using a rigorous strategy for identifying available data and strict criteria for determination of methodological quality. The consistency of thromboelastometry parameters is pointed out in **Chapter 3**. In a cross-sectional analysis the consistency of thromboelastometry parameters was analysed systematically within individual tests (INTEM, EXTEM, FIBTEM, HEPTM, APTM), between different analysers, between different channels of the same analyser, between morning and afternoon measurements (circadian variation) and if measured four weeks apart. Major factors of influence on thromboelastometry measurements are demonstrated in **Chapter 4**. In an in-vitro investigation it was investigated, to what extent parameters depend on haematocrit level and platelet count. The appendix of Chapter 4 discusses the significance of platelet count as a confounding variable in a study employing thromboelastometry measurements in patients with myeloproliferative disorders. **Chapter 5** reports about the accuracy of point-of-care coagulometers if used by patients determining their INR level themselves. Not only laboratory tests are used for diagnostic work-up of bleeding and thromboembolic disorders, but also clinical assessment tools. **Chapter 6** describes the reliability of the 4T's score in patients with suspected heparin-induced thrombocytopenia. Furthermore, the appendix of Chapter 6 discusses the value of the 4T's score in critically ill patients in view of a large diagnostic accuracy study. **Chapter 7** covers the variability of coagulation test measurements between different laboratories. The degree of, as well as the sources of variation between laboratories performing tests with identical platforms, were investigated. **Chapter 8** displays the methodological quality and extent of bias in studies investigating diagnostic tests for venous thromboembolic diseases. A systematic literature search was done and sample sizes as well as application and results of quality assessment tools extracted. Furthermore, a meta-analysis compared the pooled diagnostic odds ratios of studies unlikely to be biased with studies likely to be biased according to the QUADAS-2 quality items. **Chapter 9** contains an overall summary of this thesis and provides interpretations and future directions in this fascinating clinical field of haemostasis.

References

1. Furie B, Furie BC (2008) Mechanisms of thrombus formation. *N Engl J Med* 359: 938-949.
2. (2010) Global burden of disease. Geneva (Switzerland): World Health Organization (WHO).
3. Murray CJ, Lopez AD (1997) Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 349: 1498-1504.
4. Teixeira PG, Inaba K, Hadjizacharia P, Brown C, Salim A, et al. (2007) Preventable or potentially preventable mortality at a mature trauma center. *J Trauma* 63: 1338-1346; discussion 1346-1337.
5. (1999-2007) WISQARS Leading Causes of Death Reports. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.
6. Dossett LA, Riesel JN, Griffin MR, Cotton BA (2011) Prevalence and implications of preinjury warfarin use: an analysis of the National Trauma Databank. *Arch Surg* 146: 565-570.
7. Weinberger J (2005) Adverse effects and drug interactions of antithrombotic agents used in prevention of ischaemic stroke. *Drugs* 65: 461-471.
8. Delaney JA, Opatrny L, Brophy JM, Suissa S (2007) Drug drug interactions between antithrombotic medications and the risk of gastrointestinal bleeding. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* 177: 347-351.
9. Guh S, Grosse SD, McAlister S, Kessler CM, Soucie JM (2012) Health care expenditures for Medicaid-covered males with haemophilia in the United States, 2008. *Haemophilia* 18: 276-283.
10. Bagot CN, Arya R (2008) Virchow and his triad: a question of attribution. *Br J Haematol* 143: 180-190.
11. Mackman N (2012) New insights into the mechanisms of venous thrombosis. *J Clin Invest* 122: 2331-2336.
12. Reitsma PH, Versteeg HH, Middeldorp S (2012) Mechanistic view of risk factors for venous thromboembolism. *Arterioscler Thromb Vasc Biol* 32: 563-568.
13. Borissoff JI, Spronk HM, ten Cate H (2011) The hemostatic system as a modulator of atherosclerosis. *N Engl J Med* 364: 1746-1760.
14. Bounameaux H, Perrier A, Righini M (2010) Diagnosis of venous thromboembolism: an update. *Vasc Med* 15: 399-406.
15. Bates SM, Jaeschke R, Stevens SM, Goodacre S, Wells PS, et al. (2012) Diagnosis of DVT: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e351S-418S.
16. Eichinger S, Heinze G, Jandeck LM, Kyrle PA (2010) Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. *Circulation* 121: 1630-1636.
17. Kyrle PA, Rosendaal FR, Eichinger S (2010) Risk assessment for recurrent venous thrombosis. *Lancet* 376: 2032-2039.
18. Garcia DA, Baglin TP, Weitz JI, Samama MM (2012) Parenteral anticoagulants: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e24S-43S.
19. Ageno W, Gallus AS, Wittkowsky A, Crowther M, Hylek EM, et al. (2012) Oral anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e44S-88S.
20. Lijfering WM, Rosendaal FR, Cannegieter SC (2010) Risk factors for venous thrombosis-current understanding from an epidemiological point of view. *Br J Haematol* 149: 824-833.
21. Kouides PA (2008) Bleeding symptom assessment and hemostasis evaluation of menorrhagia. *Curr Opin Hematol* 15: 465-472.
22. Frith D, Davenport R, Brohi K (2012) Acute traumatic coagulopathy. *Current opinion in anaesthesiology* 25: 229-234.

23. Greinacher A, Althaus K, Krauel K, Selleng S (2010) Heparin-induced thrombocytopenia. *Hamostaseologie* 30: 17-28.
24. Aradi D, Komocsi A, Vorobcsuk A, Rideg O, Tokes-Fuzesi M, et al. (2010) Prognostic significance of high on-clopidogrel platelet reactivity after percutaneous coronary intervention: systematic review and meta-analysis. *Am Heart J* 160: 543-551.
25. Bates SM, Weitz JI (2005) Coagulation assays. *Circulation* 112: e53-60.
26. Goldhaber SZ, Bounameaux H (2012) Pulmonary embolism and deep vein thrombosis. *Lancet* 379: 1835-1846.
27. Anthony CJ, Karim S, Ackroyd-Stolarz S, Fry A, Murphy NG, et al. (2011) Intensity of anticoagulation with warfarin and risk of adverse events in patients presenting to the emergency department. *The Annals of pharmacotherapy* 45: 881-887.
28. Heit JA, Lahr BD, Petterson TM, Bailey KR, Ashrani AA, et al. (2011) Heparin and warfarin anticoagulation intensity as predictors of recurrence after deep vein thrombosis or pulmonary embolism: a population-based cohort study. *Blood* 118: 4992-4999.
29. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, et al. (2007) Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med* 167: 239-245.
30. Afshari A, Wikkelsø A, Brok J, Møller AM, Wetterslev J (2011) Thrombelastography (TEG) or thromboelastometry (ROTEM) to monitor haemotherapy versus usual care in patients with massive transfusion. *Cochrane Database Syst Rev*: CD007871.
31. Cuker A (2011) Heparin-induced thrombocytopenia (HIT) in 2011: An epidemic of overdiagnosis. *Thrombosis and haemostasis* 106: 993-994.
32. Cuker A, Ptashkin B, Konkole BA, Pipe SW, Whinna HC, et al. (2009) Interlaboratory agreement in the monitoring of unfractionated heparin using the anti-factor Xa-correlated activated partial thromboplastin time. *J Thromb Haemost* 7: 80-86.
33. Cuker A, Raby A, Moffat KA, Flynn G, Crowther MA (2010) Interlaboratory variation in heparin monitoring: Lessons from the Quality Management Program of Ontario coagulation surveys. *Thromb Haemost* 104: 837-844.
34. Banfi G, Del Fabbro M (2009) Biological variation in tests of hemostasis. *Semin Thromb Hemost* 35: 119-126.
35. Rosenson RS, Mosca L, Staffileno BA, Tangney CC (2001) Variability in fibrinogen measurements: an obstacle to cardiovascular risk stratification. *Atherosclerosis* 159: 225-230.
36. Chitlur M, Lusher J (2010) Standardization of thromboelastography: values and challenges. *Semin Thromb Hemost* 36: 707-711.
37. Frith D, Brohi K (2010) The acute coagulopathy of trauma shock: clinical relevance. *Surgeon* 8: 159-163.

CHAPTER 2

Technical efficacy of thromboelastometry and thrombelastography analysis in the evaluation of haemostasis: A Systematic Review

Michael Nagler[#], Sandra Ivic[#], Stefanie Nil, Hugo ten Cate, Lucas M. Bachmann, Walter A. Willemin ([#]contributed equally)

Submitted for publication

Abstract

Background: Assessing the technical efficacy of a new test against established methods among healthy volunteers should precede clinical introduction to avoid that premature application harms patients. To what extent this approach has been adopted in the assessment of thromboelastometry (ROTEM) and thrombelastography (TEG) analysis, has not been investigated thoroughly.

Objective: To systematically retrieve and describe published reports targeting the correlation of ROTEM and TEG measurements with established parameters of haemostasis in healthy subjects.

Methods: Systematic Review: Papers were searched in Medline, Scopus and the Science Citation Index database. Reference lists of included studies and of relevant reviews were screened. To be included papers had to assess technical efficacy among healthy volunteers. Two reviewers screened papers for inclusion, read full texts of potentially relevant papers and extracted data of included papers.

Results: Searches identified 1531 records of which 1527 were either excluded immediately or after reading the full text. The remaining 4 studies enrolled 243 subjects. All papers assessed TEG. Three papers provided data on the association with established parameters of haemostasis but only one study reported correlation parameters properly. Overall correlation was limited and ranged from 0.21 (coagulation index vs. activated partial thromboplastin time [aPTT] and time to maximal thrombus generation vs. aPTT) to 0.40 (total thrombus generation vs. fibrinogen level).

Conclusions: Evidence on the technical efficacy of ROTEM and TEG is sparse. The relationship between measurements of ROTEM or TEG and established methods of haemostasis remains ill understood and appears to be limited if known. In view of the scarcity of data further studies providing compelling results are required to justify more upstream evaluations in specific clinical situations.

Introduction

Back in 1978 Loop and Lusted proposed the first phased evaluation of new medical tests [1]. To date, after the publication of at least eighteen additional recommendations, (reviewed in [2]) there is a broad consensus that first phase evaluations should address parameters of technical efficacy such as minimal detection level, circadian fluctuation, reproducibility and correlation with established tests in healthy volunteers to avoid that premature application in clinical practice leads to wrong diagnoses, unnecessary delays in starting appropriate treatment or even hazardous therapy. It has been recognised that sophisticated and expensive tests that are disseminated without suitable evaluations can subsequently be found to have marginal clinical value and economic benefit. Well known examples include the carcinoembryonic antigen test in the diagnosis of colon cancer [3], Iodine 125-labeled fibrinogen scans in the diagnosis of deep venous thrombi [4] or rapid magnetic resonance imaging in the management of patients with low back pain [5].

Thromboelastometry (ROTEM) and thrombelastography (TEG) analysis, two novel methods to evaluate haemostasis, are increasingly used in practice based on the expectation by clinicians that the implementation of these tests will substantially improve the management of acute bleeding. Several advantages over established laboratory parameters promote their use in emergency unit and operation theatre. The analyser is easy to use, it works fast, and the tests are consistently available. The graphical presentation of the results suggests intuitively an interpretation [6]. Furthermore, tests claim to provide global information on all aspects of haemostasis including fibrinogen level, platelet function, coagulation cascade, cross-linking of fibrin and fibrinolysis. In addition, whereas the therapeutic consequences of results obtained with established parameters often remains uncertain, ROTEM/TEG results can translate immediately into specific interventions to enhance haemostasis. Several studies explored their role in the detection of coagulopathies and changes in bleeding management as well as in the perioperative setting [7-12]. However, despite a vast amount of investigations, the clinical value of these methods remains largely left in the dark [13-17]. The supposed lack of rigorous examinations of their technical efficiency, i.e. the correlation of ROTEM and TEG measurements with established parameters of haemostasis among healthy individuals, is seen as the key concern in several recent papers [6,18,19].

In the absence of a systematic review of the available literature, it remains difficult to assess the technical qualities of TEG and ROTEM applications. We therefore performed a systematic review to search, describe and to draw up an inventory of published reports targeting at the correlation of ROTEM and TEG measurements with established parameters of haemostasis among healthy subjects, applying rigorous reviewing methods.

Methods

These systematic reviews were done according to the procedure mentioned below and the PRISMA statement [20]. A formal, registered study protocol was not implemented.

Search strategy

To identify papers investigating the technical efficacy of ROTEM and TEG devices we searched (PRE-)MEDLINE (Pubmed interface) using a search string including the two Medical Subject Headings (MeSH) “Thrombelastography/methods”[Mesh], “Blood Coagulation Tests”[Mesh] and the free text terms thromboelastography, thromboelastometry, ROTEM and TEG. SCOPUS was searched using the following algorithm: TITLE-ABS-KEY(thromboelastography OR thromboelastometry)) AND (valid*) AND (LIMIT-TO(DOCTYPE, “ar”)) AND (LIMIT-TO(SUBJAREA, “MEDI”)) We also searched the Web of Science database entering 4 seminal papers [6,21-23] and checked their citations for potentially relevant papers. Searches were complemented checking reference lists of included studies and of a recently published Cochrane review [13]. Electronic searches were done from inception to July 2012. An update search of (PRE-)MEDLINE was performed in June 2013.

Selection Criteria for Inclusion

To be included a paper had to use whole blood or citrated blood, blood from healthy volunteers and had to assess either ROTEM or TEG in combination with at least one established parameter of haemostasis such as activated partial thromboplastin time (aPTT), prothrombin time (PT), platelet count (PLT), fibrinogen level, or d-dimers. We excluded papers assessing trauma patients, patients undergoing cardiac surgery, or patients with sepsis and other patients with severe illness and papers without a comparison to an established method of haemostasis. This also included head-to-head comparisons between ROTEM and TEG if they were assessed in absence of an established haemostasis method.

Selection process and data abstraction

In the case of multiple publications from the same participants, the most complete report was chosen for each study. We extracted data in duplicate and a third reviewer resolved any discrepancies if the two reviewers disagreed. Of each study we extracted participants' age, number of female participants, the total number of included subjects, the device employed (ROTEM, TEG), which of the parameters was assessed (R time, K time, α , MA (maximal amplitude), SEMS (shear elastic modulus strength), CI (coagulation index), MTG (maximal thrombus generation), TMG (time to maximal thrombus generation), TTG (total thrombus generation)). Finally, we also extracted which of the

established methods was used as the comparator tests and what type of correlation or concordance parameters was used.

Results

Searches retrieved 1531 records out of which 1407 had to be excluded after screening title or abstract. Full texts of 124 articles were examined for inclusion. 118 articles fulfilled at least one exclusion criterion leaving 4 articles for analysis [24-27]. All papers assessed TEG. The selection process is described in **Figure 1**.

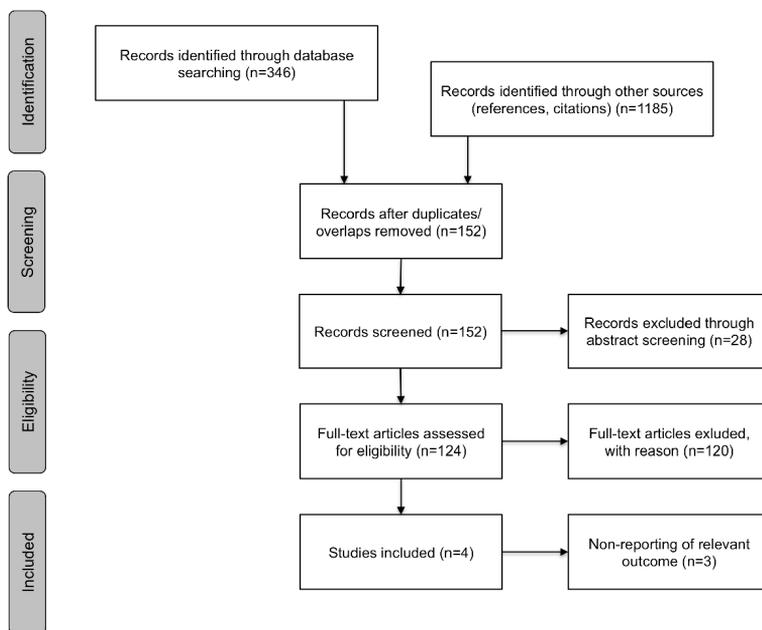


Figure 1: Study flow

Of these four studies only three reported the extent of association between pairs of parameters. Two studies reported coefficients from regression analyses and only one study quantified the correlation properly using correlation coefficients. Summaries of included studies are shown in **Tables 1 and 2**.

Table 1: Baseline characteristics

| Author | Year | Mean age | SD mean age | Age range | Female | Male | Total persons | Device | Correlation |
|------------|------|----------|-------------|-----------|--------|------|---------------|--------|--|
| Roeloffzen | 2010 | 49.50 | 25.46 | 19-87 | 60 | 60 | 120 | TEG | Pearson correlation coefficients (r^2) |
| Foley | 2012 | | | 29-32 | 0 | 5 | 5 | TEG | - |
| Scarpelini | 2009 | 36.80 | 11.20 | | 40 | 78 | 118 | TEG | Linear regression |
| Rivard | 2005 | | | | 2 | 2 | 4 | TEG | Linear and non-linear regression |

In a study enrolling 120 participants (50% female) of 49.5 years on average, Roeloffzen and colleagues compared R time with the activated partial thromboplastin time (aPTT), K time with the haemoglobin level, aPTT and the fibrinogen level, α with aPTT and the fibrinogen level, the maximal amplitude (MA) with the fibrinogen level, the shear elastic modulus strength (SEMS) with the fibrinogen level, the coagulation index (CI) with aPTT and the fibrinogen level, the maximal thrombus generation (MTG) with the fibrinogen level, the time to maximal thrombus generation (TMG) with aPTT and the total thrombus generation (TTG) with the fibrinogen level.

Overall correlation was limited and ranged from 0.21 (coagulation index vs. activated partial thromboplastin time [aPTT] and time to maximal thrombus generation vs. aPTT) to 0.40 (total thrombus generation vs. fibrinogen level).

Table 2: Data extraction

| Author | Year | Population size | Device | Test | Hemoglobin level [Hb] | Platelet count | Activated partial thromboplastin time [aPTT] | Fibrinogen level | Thrombin/antithrombin complex [TAT] |
|------------|------|-----------------|--------|---|-----------------------|----------------|--|------------------|-------------------------------------|
| Roeloffzen | 2010 | 120 | TEG | <i>R</i> time | NS | NS | 0.36 | NS | - |
| Roeloffzen | 2010 | 120 | TEG | <i>K</i> time | 0.24 | NS | 0.24 | -0.27 | - |
| Roeloffzen | 2010 | 120 | TEG | α | NS | NS | -0.22 | 0.28 | - |
| Roeloffzen | 2010 | 120 | TEG | MA (maximal amplitude) | NS | NS | NS | 0.37 | - |
| Roeloffzen | 2010 | 120 | TEG | SEMS (shear elastic modulus strength) | NS | NS | NS | 0.37 | - |
| Roeloffzen | 2010 | 120 | TEG | CI (coagulation index) | NS | NS | -0.21 | 0.25 | - |
| Roeloffzen | 2010 | 120 | TEG | MTG (maximal thrombus generation) | NS | NS | NS | 0.27 | - |
| Roeloffzen | 2010 | 120 | TEG | TMG (time to maximal thrombus generation) | NS | NS | 0.21 | NS | - |
| Roeloffzen | 2010 | 120 | TEG | TTG (total thrombus generation) | NS | NS | NS | 0.40 | - |
| Rivard | 2005 | 4 | TEG | TTG (total thrombus generation) | - | - | - | - | 0.85* |
| Rivard | 2005 | 4 | TEG | TTG (total thrombus generation) | - | - | - | - | 0.94* |

NS = Not Stated, * = derived from regression analyses

Discussion

Using comprehensive retrieval methods, our systematic review identified no study of ROTEM and only four studies of TEG investigating technical efficacy in healthy volunteers. One single study reported correlations properly and revealed only a weak to moderate correlation with established parameters of haemostasis.

There are several studies published investigating the technical efficacy ROTEM or TEG in animals, as for example Mauch et al [28]. In this study the intrarater and inter-rater validity of the ROTEM delta was analysed with blood derived from healthy pigs. Another prospective study compared the coagulation profiles of healthy foals using standard techniques (prothrombin time (PT), activated partial thromboplastin time (aPTT), fibrinogen concentration (Fib) and antithrombin (AT)) with the TEG analyser [29]. In humans, the validity of the ROTEM and TEG was verified through comparison with standard methods in the context of specific circumstances, as for example orthotopic liver transplantation [30-32], during aortic surgery [12,33,34] or in trauma patients [16,35,36]. We also found a trial reporting on head to head comparison of ROTEM and TEG [19]. Although these approaches also yield valuable information, we think that a first comparison with standardized methods in healthy subjects would have been required in order to evaluate the reproducibility of the obtained results.

We applied up-to-date and rigorous systematic review methods to retrieve and assess the available evidence. In view of the risk that potentially relevant studies could be missed due to ambiguous indexing in the various databases, we applied an over-inclusive approach. These sensitive searches retrieved over 1500 records. Screening for inclusion was made in duplicate to reduce the risk of missing relevant papers. We expected to find a reasonable number of studies that would allow us to provide quantitative summaries of various comparisons. Moreover, out of the many available, we hoped to depict those parameters from thromboelastometry and thrombelastography that would be most suited for investigations in various clinical settings such as acute bleeding, applications in emergency rooms and of patients with specific haemostatic deficiencies. Despite our motivation the current evidence base impeded us from generating meaningful summaries for further research and clinical practice.

The lack of studies examining the ROTEM or TEG device in a cohort of healthy volunteers and comparing its findings to established methods of haemostasis systematically is remarkable. Despite the absence of that fundamental research, these devices are already in daily use and no one can foresee possible consequences of premature implementation into clinical practice. Furthermore, it cannot be ruled out that decisions made on the basis of miss-interpretations, false alarms or false negative findings can increase patients' burden.

One may argue, that the clinical value of established coagulation parameters is limited as well. It is true, that results of coagulation parameters such as PT not directly prompt any specific treatment. However, these parameters remain the gold standard for detection of coagulopathies in the setting of acute bleeding [37] and treatment based on ROTEM or TEG assays didn't show any relevant clinical benefits so far [13,17].

Following existing guidelines of test evaluation (reviewed and discussed in: [2]) we propose that further research should first investigate the distribution of normal values of ROTEM and TEG parameters in reasonably sized studies enrolling healthy subjects. Second, distributions should be assessed in groups of patients with a specific illness and compared to distributions of healthy subjects. In a third step, assessing test performance characteristics such as sensitivity and specificity should be performed, taking contextual clinical information into account. Fourth, studies should determine the marginal informativeness of ROTEM and TEG in context of other clinical information about the patients' state that is available prior to testing. Finally, formal decision-analytic models should be performed determining the optimal areas for clinical use taking patient preferences and clinical outcomes into account.

In conclusion, the relationship between measurements of ROTEM or TEG and established methods of haemostasis remains ill understood and appears to be extremely limited. In view of the scarcity of data further studies providing compelling results are required to justify more upstream evaluations in specific clinical situations.

References

1. Loop JW, Lusted LE (1978) American College of Radiology Diagnostic Efficacy Studies. *AJR Am J Roentgenol* 131: 173-179.
2. Lijmer JG, Leeflang M, Bossuyt PM (2009) Proposals for a phased evaluation of medical tests. *Med Decis Making* 29: E13-21.
3. Fletcher RH (1986) Carcinoembryonic antigen. *Ann Intern Med* 104: 66-73.
4. Lensing AW, Hirsh J (1993) 125I-fibrinogen leg scanning: reassessment of its role for the diagnosis of venous thrombosis in post-operative patients. *Thromb Haemost* 69: 2-7.
5. Hollingworth W, Gray DT, Martin BI, Sullivan SD, Deyo RA, et al. (2003) Rapid magnetic resonance imaging for diagnosing cancer-related low back pain. *J Gen Intern Med* 18: 303-312.
6. Luddington RJ (2005) Thrombelastography/thromboelastometry. *Clin Lab Haematol* 27: 81-90.
7. Fenger-Eriksen C, Ingerslev J, Tonnesen E, Sorensen B (2009) Citrate artificially masks the haemostatic effect of recombinant factor VIIa in dilutional coagulopathy. *Ann Hematol* 88: 255-260.
8. Fenger-Eriksen C, Jensen TM, Kristensen BS, Jensen KM, Tonnesen E, et al. (2009) Fibrinogen substitution improves whole blood clot firmness after dilution with hydroxyethyl starch in bleeding patients undergoing radical cystectomy: a randomized, placebo-controlled clinical trial. *J Thromb Haemost* 7: 795-802.
9. Girdauskas E, Kempfert J, Kuntze T, Borger MA, Enders J, et al. Thromboelastometrically guided transfusion protocol during aortic surgery with circulatory arrest: a prospective, randomized trial. *J Thorac Cardiovasc Surg* 140: 1117-1124 e1112.
10. Rugeri L, Levrat A, David JS, Delecroix E, Floccard B, et al. (2007) Diagnosis of early coagulation abnormalities in trauma patients by rotation thrombelastography. *J Thromb Haemost* 5: 289-295.
11. Shore-Lesserson L, Manspeizer HE, DePerio M, Francis S, Vela-Cantos F, et al. (1999) Thromboelastography-guided transfusion algorithm reduces transfusions in complex cardiac surgery. *Anesth Analg* 88: 312-319.
12. Girdauskas E, Kempfert J, Kuntze T, Borger MA, Enders J, et al. (2010) Thromboelastometrically guided transfusion protocol during aortic surgery with circulatory arrest: a prospective, randomized trial. *J Thorac Cardiovasc Surg* 140: 1117-1124 e1112.
13. Afshari A, Wikkelsø A, Brok J, Møller AM, Wetterslev J (2011) Thrombelastography (TEG) or thromboelastometry (ROTEM) to monitor haemotherapy versus usual care in patients with massive transfusion. *Cochrane Database Syst Rev*: CD007871.
14. Coakley M, Hall JE, Evans C, Duff E, Billing V, et al. (2011) Assessment of thrombin generation measured before and after cardiopulmonary bypass surgery and its association with postoperative bleeding. *J Thromb Haemost* 9: 282-292.
15. Lee GC, Kicza AM, Liu KY, Nyman CB, Kaufman RM, et al. (2012) Does rotational thromboelastometry (ROTEM) improve prediction of bleeding after cardiac surgery? *Anesthesia and analgesia* 115: 499-506.
16. Raza I, Davenport R, Rourke C, Platton S, Manson J, et al. (2012) The incidence and magnitude of fibrinolytic activation in trauma patients. *J Thromb Haemost*.
17. Wikkelsøe AJ, Afshari A, Wetterslev J, Brok J, Moeller AM (2011) Monitoring patients at risk of massive transfusion with Thrombelastography or Thromboelastometry: a systematic review. *Acta anaesthesiologica Scandinavica* 55: 1174-1189.
18. Chitlur M, Lusher J (2010) Standardization of thromboelastography: values and challenges. *Semin Thromb Hemost* 36: 707-711.
19. Chitlur M, Sorensen B, Rivard GE, Young G, Ingerslev J, et al. (2011) Standardization of thromboelastography: a report from the TEG-ROTEM working group. *Haemophilia* 17: 532-537.
20. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med* 3: e123-130.

21. Di Benedetto P, Baciarello M, Cabetti L, Martucci M, Chiaschi A, et al. (2003) Thrombelastography. Present and future perspectives in clinical practice. *Minerva Anesthesiol* 69: 501-509, 509-515.
22. Lang T, Bauters A, Braun SL, Potzsch B, von Pape KW, et al. (2005) Multi-centre investigation on reference ranges for ROTEM thromboelastometry. *Blood Coagul Fibrinolysis* 16: 301-310.
23. Salooja N, Perry DJ (2001) Thrombelastography. *Blood Coagul Fibrinolysis* 12: 327-337.
24. Foley JH, Butenas S, Mann KG, Brummel-Ziedins KE (2012) Measuring the mechanical properties of blood clots formed via the tissue factor pathway of coagulation. *Anal Biochem* 422: 46-51.
25. Rivard GE, Brummel-Ziedins KE, Mann KG, Fan L, Hofer A, et al. (2005) Evaluation of the profile of thrombin generation during the process of whole blood clotting as assessed by thrombelastography. *J Thromb Haemost* 3: 2039-2043.
26. Roeloffzen WW, Kluin-Nelemans HC, Mulder AB, Veeger NJ, Bosman L, et al. (2010) In normal controls, both age and gender affect coagulability as measured by thrombelastography. *Anesth Analg* 110: 987-994.
27. Scarpelini S, Rhind SG, Nascimento B, Tien H, Shek PN, et al. (2009) Normal range values for thromboelastography in healthy adult volunteers. *Braz J Med Biol Res* 42: 1210-1217.
28. Mauch J, Spielmann N, Hartnack S, Madjdpour C, Kutter AP, et al. (2011) Intrarater and interrater variability of point of care coagulation testing using the ROTEM delta. *Blood Coagul Fibrinolysis* 22: 662-666.
29. Mendez-Angulo JL, Mudge M, Zaldivar-Lopez S, Vilar-Saavedra P, Couto G (2011) Thromboelastography in healthy, sick non-septic and septic neonatal foals. *Aust Vet J* 89: 500-505.
30. Coakley M, Reddy K, Mackie I, Mallett S (2006) Transfusion triggers in orthotopic liver transplantation: a comparison of the thromboelastometry analyzer, the thromboelastogram, and conventional coagulation tests. *J Cardiothorac Vasc Anesth* 20: 548-553.
31. Herbstreit F, Winter EM, Peters J, Hartmann M (2010) Monitoring of haemostasis in liver transplantation: comparison of laboratory based and point of care tests. *Anaesthesia* 65: 44-49.
32. Roullet S, Pillot J, Freyburger G, Biais M, Quinart A, et al. (2010) Rotation thromboelastometry detects thrombocytopenia and hypofibrinogenaemia during orthotopic liver transplantation. *Br J Anaesth* 104: 422-428.
33. Ogawa S, Szlam F, Chen EP, Nishimura T, Kim H, et al. (2012) A comparative evaluation of rotation thromboelastometry and standard coagulation tests in hemodilution-induced coagulation changes after cardiac surgery. *Transfusion* 52: 14-22.
34. Solomon C, Cadamuro J, Ziegler B, Schochl H, Varvenne M, et al. (2011) A comparison of fibrinogen measurement methods with fibrin clot elasticity assessed by thromboelastometry, before and after administration of fibrinogen concentrate in cardiac surgery patients. *Transfusion* 51: 1695-1706.
35. Davenport R, Manson J, De'Ath H, Platton S, Coates A, et al. (2011) Functional definition and characterization of acute traumatic coagulopathy. *Crit Care Med* 39: 2652-2658.
36. Tauber H, Innerhofer P, Breitkopf R, Westermann I, Beer R, et al. (2011) Prevalence and impact of abnormal ROTEM(R) assays in severe blunt trauma: results of the 'Diagnosis and Treatment of Trauma-Induced Coagulopathy (DIA-TRE-TIC) study'. *Br J Anaesth* 107: 378-387.
37. Frith D, Goslings JC, Gaarder C, Maegele M, Cohen MJ, et al. (2010) Definition and drivers of acute traumatic coagulopathy: clinical and experimental investigations. *Journal of thrombosis and haemostasis: JTH* 8: 1919-1925.

CHAPTER 3

Consistency of thromboelastometry analysis under scrutiny: Results of a systematic evaluation within and between analysers

Michael Nagler, Hugo ten Cate, Silvio Kathriner, Mattias Casutt, Lucas M. Bachmann, Walter A. Wuillemin

Thrombosis and Haemostasis 2014, 111(6): 1161-66

Abstract

Background: While the use of thromboelastometry analysis (ROTEM®) in evaluation of haemostasis is rapidly increasing, important validity parameters of testing remain inadequately examined. We aimed to study systematically the consistency of thromboelastometry parameters within individual tests regarding measurements between different analysers, between different channels of the same analyser, between morning and afternoon measurements (circadian variation), and if measured four weeks apart.

Methods: Citrated whole blood samples from 40 healthy volunteers were analysed with two analysers in parallel. EXTEM, INTEM, FIBTEM, HEPTEM and APTEM tests were conducted. A Bland-Altman comparison was performed and homogeneity of variances was tested using the pitman test. P-value ranges were used to classify the level of homogeneity ($p < 0.15$ – low homogeneity, $p = 0.15$ to 0.5 – intermediate homogeneity, $p > 0.5$ high homogeneity).

Results: Less than half of all comparisons made showed high homogeneity of variances ($p > 0.5$) and in about a fifth of comparisons data distributions were heterogeneous ($p < 0.15$). There was no clear pattern for homogeneity. On average, comparisons of MCF, ML and LI30 measurements tended to be better, but none of the tests assessed outperformed another.

Conclusions: Systematic investigation reveals large differences in the results of some thromboelastometry parameters and lack of consistency. Clinicians and scientists should take these inconsistencies into account and focus on parameters with a higher homogeneity such as MCF.

Introduction

To be useful in haemostasis assessment, thromboelastometry analyses (ROTEM®) should not depend on the device itself, or on the moment of testing, but need to reflect the haemostatic condition of a patient. Thromboelastometry analysis (ROTEM®) is a point-of-care viscoelastic test comparable to thromboelastography (TEG®), which has been enthusiastically received due to its convenient handling features by many clinicians. Reasons for the successful introduction of the ROTEM® device include good handle ability for non-laboratory personnel, the quick delivery and graphical display of results. All these features facilitate its use in emergency rooms and operation theatres [1,2]. Moreover, various authors suggested performing thromboelastometry in the setting of acute bleeding because it is claimed that this provides global information on all parts of haemostasis including fibrinogen levels, platelet function, coagulation cascade, cross-linking of fibrin and fibrinolysis [3,4]. Thromboelastometry has been evaluated mainly in perioperative and trauma settings for detection of coagulopathy, guiding substitution therapy and for reducing transfusions [3,5-13]. Furthermore, thromboelastometry is mentioned in several recent guidelines of scientific societies [14,15]. However, although a vast amount of investigations have been performed in the last few years, the clinical value remains still unclear [16-21]. Surprisingly, important validity parameters of testing remain incompletely examined and even key questions on consistency are still inadequately determined [22,23]. Limitations in reproducibility may seriously reduce the value of a medical test, possibly lead to wrong diagnoses, unnecessary delays in starting appropriate treatment, or even hazardous therapy. Thus, evidence-based laboratory guidelines recommend performing these investigations in the early evaluation of a medical test [24,25].

Previous investigations studying reliability, revealed inconsistent data and left one essential question open: Are thromboelastometry measurements reproducible in clinically relevant situations [4,6,22,23,26-30]? Investigations focussed on partial aspects only, surrogate test materials were used instead of whole blood samples, sample sizes were small, and the distinct characteristics of the analysis that limit the application of traditional measures of reliability such as within-run or inter-assay imprecision have not been respected. The whole blood samples are vulnerable to artefacts due to ex vivo activation of thrombocytes, coagulation factors and inhibitors; the analysers have a multi-channel composition; and the duration of a single measurement is very long. In addition, key questions of consistency such as reproducibility of measurements between different analysers and intra-individual variation have not been adequately addressed. Furthermore, every single test run produces a large number of results parameters (at least 25), and no consensus exists which of them should be taken into account. Several recent reviews and other publications discussed possible problems in reproduc-

ibility and argued that efforts are necessary to establish reliability and standardisation [2,22,26]. Therefore, the aim of the present investigation was to study the consistency of thromboelastometry parameters within individual tests, between different analysers, between different channels of the same analyser, between morning and afternoon measurements (circadian variation), and if measured four weeks apart (day-to-day variation) systematically in a joint assessment.

Materials and Methods

Study design

In this prospective study citrated whole blood samples were taken from 40 healthy young volunteers and assessed with two ROTEM analysers in parallel. Individuals did not have any known disease, were taking no medication with an effect on haemostatic parameters and had no history of bleeding or thromboembolic events. The study was approved by the local ethical review board (Kantonale Ethikkommission Luzern; #11001) and all participants provided written informed consent.

Sampling and preanalytic procedures

Blood withdrawal was done under standardised conditions to preclude preanalytic influences. Plastic syringes (Monovette[®], Sarstedt, Nümbrecht, Germany) containing 1 ml trisodium citrate (0.106 mol/l) for 9 ml of blood were used for collection of whole blood samples by direct venipuncture of the median cubital vein. No heparin-containing peripheral or central venous catheter was used. The citrated sample for thromboelastometry was collected after discarding the first syringe. Platelet count and haematocrit were determined using XE 5000 (Sysmex AG, Horgen, Switzerland). Time between blood withdrawal and analysis was strictly held at 10 to 15 minutes to eliminate possible impact of platelet activation, activation of coagulation factors and inhibitors. This time period included 5 minutes of pre-heating to 37°C to achieve an equal temperature of all samples. To avoid possible influences due to different preparation and handling, the same laboratory technician performed all tests.

Thromboelastometry analysis

Thromboelastometry was conducted using two new ROTEM[®] analysers (ROTEM[®] delta; Pentapharm GmbH, Munich, Germany). Thromboelastometry procedures were performed according to the manufacturer's instructions by following an automated electronic pipette programme. Details of the ROTEM[®] analysis, parameters and activating agents have been described in detail elsewhere [5,26,31,32]. In brief, thromboelastometry analysis assesses the viscoelastic properties of a forming clot by tracing the oscillations of a pin within a cup containing a citrated whole blood sample. Different aspects

of clot formation are recorded as *parameters* of the analysis: time from addition of the activator to start of clotting is measured as clotting time (CT, s); the acceleration of clotting is represented by the clot formation time (CFT, s), and the alpha angle (ALP, degree); the clot strength is indicated by the maximum clot firmness (MCF, mm), amplitude achieved after 10 minutes (A10, mm), after 20 minutes (A20, mm), and after 30 minutes (A30, mm); and the amount of fibrinolysis is demonstrated by the ratio of clot strength after 60 minutes (ML, %), and 30 minutes (LI30, %) to the MCF value. The different *tests* of the analysis represent the clotting response after addition of several activators. INTEM monitors the intrinsic pathway of coagulation by the use of a respective reagent. A tissue factor-containing reagent is employed in EXTEM, representing the extrinsic pathway. In FIBTEM, a platelet inhibitor is added demonstrating the (functional) fibrinogen part of clotting. The fibrinolysis blocking agent aprotinin is added in APTEM, allowing to detect hyperfibrinolysis. In HEPTEM, heparinase is added allowing detection of heparin in INTEM.

Reproducibility between different ROTEM analysers

INTEM, EXTEM and FIBTEM parameters were assessed in parallel with two ROTEM[®] analysers in three channels using whole blood samples of 20 individuals. The fourth channel was used to determine APTEM parameters in 10 samples and HEPTEM in 10 samples, each.

Reproducibility between different channels of the same ROTEM analyser

EXTEM and INTEM parameters were measured in channels 1+2 and 3+4, respectively, using two ROTEM[®] analysers, resulting in 40 measurements of 20 individuals corresponding to 20 measurements in parallel.

Intra-individual circadian variation between morning and afternoon measurements ROTEM[®] analysis was conducted in 20 individuals at 8:00 a.m. and 4:00 p.m. of the same day using the same ROTEM[®] analyser and the same channel for each test. INTEM, EXTEM and FIBTEM parameters were recorded. HEPTEM and APTEM were conducted similarly in 10 individuals.

Intra-individual variation measured four weeks apart

ROTEM[®] analysis was conducted in 20 individuals on day one as well as four weeks later using the same ROTEM[®] analyser and the same channel for each test. INTEM, EXTEM and FIBTEM parameters were recorded. HEPTEM and APTEM were conducted similarly in 10 individuals.

Statistical analysis

Differences versus means of pairs of variables were plotted (Bland-Altman plots) and the correlation between differences and means was calculated. The Pitman-test was

performed to test a null hypothesis of equal variances given bivariate normality. P-value ranges ($p < 0.15$ – low homogeneity, $p = 0.15$ to 0.5 – intermediate homogeneity, $p > 0.5$ high homogeneity) were used to classify the level of homogeneity between pairs of variances. Analyses were performed using the Stata 11.2 statistics software package. (StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.)

Results

Demographics of participants

Forty individuals were included, 20 male and 20 female. Median age was 34.5 (range 20-62), female 31.0 (21-60), male 39.5 (20-62).

Consistency within individual tests

Measurements are displayed as Bland-Altman plots for comparison between different analysers and between different channels of the same analyser (**Figure S1–supplementary material**), and for morning and afternoon measurements as well as measurements 4 weeks apart (**Figure S2**). The corresponding homogeneity of measurements is given in **Table 1**. Differences in variance varied widely within the individual tests. Some parameters show a good precision ($p > 0.5$; displayed in green), others are only moderate ($p = 0.15$ to 0.5 ; pale red) or poor ($p < 0.15$; dark red). Discrepancies were found in all situations (comparison between analysers, between different channels of the same analyser, circadian variation, and day-to-day variation).

Reproducibility between different ROTEM analysers

High homogeneity was found in 2 out of 9 measurements of EXTEM, in 4 out of 9 INTEM parameters, in FIBTEM MCF, and in 3 out of 3 APTM parameters (**Table 1, Figure S1**). A moderate homogeneity was shown in 6/9 EXTEM measurements, 3/9 INTEM parameters, and 2/3 HEPTM parameters. Poor homogeneity was found in 1/9 EXTEM measurements, 2/9 INTEM parameters, and 1/3 HEPTM parameters.

Reproducibility between different channels of the same ROTEM analyser

High homogeneity was found in 7 out of 18 measurements of EXTEM, and in 9 out of 18 INTEM parameters (**Table 1; Figure S1**). A moderate homogeneity was shown in 7/18 EXTEM measurements, and in 4/18 INTEM parameters. Poor homogeneity was found in 4/18 EXTEM measurements, and in 5/18 INTEM parameters.

Table 1: Difference in variance for thromboelastometry measurements in defined situations

| Test / Parameter | CT | CFT | ALP | A10 | A20 | MCF | ML | LI30 | A30 |
|--|---------------------------------------|---------------------------------------|--------------------------|--------------------------|--------------------------|---------------------------------------|--------------------------|--------------------------|-------------------------|
| Comparison between ROTEM analyzers | | | | | | | | | |
| EXTEM | r = 0.247, p = 0.293 | r = 0.253, p = 0.281 | r = 0.074, p = 0.757 | r = 0.421, p = 0.065 | r = 0.614, p = 0.004 | r = 0.278, p = 0.235 | r = -0.101, p = 0.674 | r = 0.250, p = 0.288 | r = 0.434, p = 0.056 |
| INTEM | r = 0.325, p = 0.162 | r = -0.527, p = 0.063 | r = -0.614, p = 0.040 | r = -0.058, p = 0.809 | r = -0.012, p = 0.959 | r = -0.115, p = 0.634 | r = 0.058, p = 0.808 | r = 0.267, p = 0.256 | r = 0.172, p = 0.468 |
| FIBTEM | | | | | | r = -0.076, p = 0.752 | | | |
| APTEM | r = -0.092, p = 0.802 ¹ | r = -0.071, p = 0.846 ¹ | | | | r = -0.105, p = 0.774 ¹ | | | |
| HEPTEM | r = 0.650, p = 0.042 ¹ | r = 0.422, p = 0.225 ¹ | | | | r = -0.265, p = 0.490 ¹ | | | |
| Comparison of circadian differences | | | | | | | | | |
| EXTEM | r = 0.407, p = 0.075 | r = 0.170, p = 0.472 | r = 0.075, p = 0.752 | r = 0.323, p = 0.165 | r = 0.385, p = 0.094 | r = 0.272, p = 0.247 | r = 0.216, p = 0.359 | r = -0.312, p = 0.223 | r = 0.340, p = 0.143 |
| INTEM | r = 0.083, p = 0.727 | r = -0.053, p = 0.823 | r = -0.156, p = 0.521 | r = 0.020, p = 0.933 | r = 0.053, p = 0.823 | r = -0.016, p = 0.946 | r = 0.147, p = 0.536 | r = -0.101, p = 0.676 | r = 0.031, p = 0.896 |
| FIBTEM | | | | | | r = -0.378, p = 0.151 | | | |
| APTEM | r = -0.226, p = 0.550 ¹ | r = 0.313, p = 0.379 ¹ | | | | r = -0.147, p = 0.685 ¹ | | | |
| HEPTEM | r = 0.616, p = 0.058 ¹ | r = 0.548, p = 0.101 ¹ | | | | r = -0.304, p = 0.435 ¹ | | | |
| Comparison after 4 weeks | | | | | | | | | |
| EXTEM | r = 0.346, p = 0.136 | r = 0.280, p = 0.232 | r = 0.260, p = 0.269 | r = 0.390, p = 0.089 | r = 0.299, p = 0.200 | r = 0.361, p = 0.118 | r = 0.020, p = 0.935 | r = 0.111, p = 0.642 | r = 0.389, p = 0.090 |
| INTEM | r = -0.266, p = 0.290 | r = -0.175, p = 0.473 | r = -0.324, p = 0.207 | r = 0.019, p = 0.937 | r = 0.217, p = 0.358 | r = 0.204, p = 0.389 | r = -0.107, p = 0.657 | r = 0.073, p = 0.760 | r = 0.166, p = 0.485 |
| FIBTEM | | | | | | r = -0.115, p = 0.634 | | | |
| APTEM | r = 0.314, p = 0.376 ¹ | r = 0.702, p = 0.024 ¹ | | | | r = 0.382, p = 0.277 ¹ | | | |
| HEPTEM | r = 0.371, p = 0.291 ¹ | r = 0.552, p = 0.098 ¹ | | | | r = -0.235, p = 0.536 ¹ | | | |
| Reproducibility between different channels of the same ROTEM analyzer (1 st device) | | | | | | | | | |
| EXTEM | r = 0.284, p = 0.225 | r = 0.649, p = 0.002 | r = 0.157, p = 0.509 | r = 0.487, p = 0.029 | r = 0.476, p = 0.034 | r = 0.199, p = 0.400 | r = -0.244, p = 0.328 | r = -0.136, p = 0.574 | r = 0.315, p = 0.177 |
| INTEM | r = -0.024, p = 0.888 | r = -0.546, p = 0.001 | r = -0.507, p = 0.001 | r = -0.199, p = 0.039 | r = -0.039, p = 0.888 | r = -0.031, p = 0.888 | r = 0.400, p = 0.001 | r = 0.120, p = 0.574 | r = 0.154, p = 0.468 |

| Test / Parameter | CT | CFT | ALP | A10 | A20 | MCF | ML | LI30 | A30 |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | p= 0.921 | p= 0.057 | p= 0.071 | p= 0.419 | p= 0.869 | p= 0.897 | p= 0.081 | p= 0.614 | p= 0.517 |
| Reproducibility between different channels of the same ROTEM analyzer (2 nd device) | | | | | | | | | |
| EXTEM | r= 0.156, p= 0.510 | r=-0.438, p= 0.106 | r=-0.080, p= 0.739 | r=-0.198, p= 0.420 | r=-0.173, p= 0.478 | r=-0.129, p= 0.595 | r=-0.094, p= 0.695 | r=-0.133, p= 0.583 | r=-0.214, p= 0.385 |
| INTEM | r=-0.182, p= 0.458 | r=-0.508, p= 0.071 | r=-0.295, p= 0.245 | r=-0.223, p= 0.367 | r=-0.033, p= 0.889 | r=-0.057, p= 0.813 | r=-0.504, p= 0.072 | r= 0.110, p= 0.646 | r=-0.010, p= 0.966 |

Difference in variance, pitman test, n = 20, ¹n = 10; p-value > 0.5: high homogeneity (green); p-value 0.15-0.5: intermediate homogeneity (pale-red); p-value < 0.15: low homogeneity (dark red)

Variation between morning and afternoon measurements

High homogeneity was found in 1 out of 9 measurements of EXTEM, in 9 out of 9 INTEM parameters, and in 2 out of 3 APTTEM parameters (**Table S1; Figure S2**). A moderate homogeneity was shown in 5/9 EXTEM measurements, in FIBTEM MCF, in 1/3 APTTEM parameters, and 1/3 HEPTTEM parameters. Poor homogeneity was found in 3/9 EXTEM measurements, and 2/3 HEPTTEM parameters.

Intra-individual variation measured four weeks apart

High homogeneity was found in 2 out of 9 measurements of EXTEM, in 3 out of 9 INTEM parameters, in FIBTEM MCF, and in 1 out of 3 HEPTTEM parameters (**Table S1; Figure S2**). A moderate homogeneity was shown in 3/9 EXTEM measurements, 6/9 INTEM parameters, and 2/3 APTTEM parameters, and 1/3 HEPTTEM parameters. Poor homogeneity was found in 4/9 EXTEM measurements, 1/3 APTTEM parameters, and 1/3 HEPTTEM parameters.

Figure S1. Bland-Altman difference plots for comparisons between (a) different analysers, and (b) different channels of the same analyser. The differences between the two measurements are plotted against the average (mean) of the two measurements (ROTEM 1, ROTEM 2 or Channel 1,2; 3,4 respectively). Horizontal lines are drawn at the mean differences and the limits of agreement (± 1.96 standard deviation).

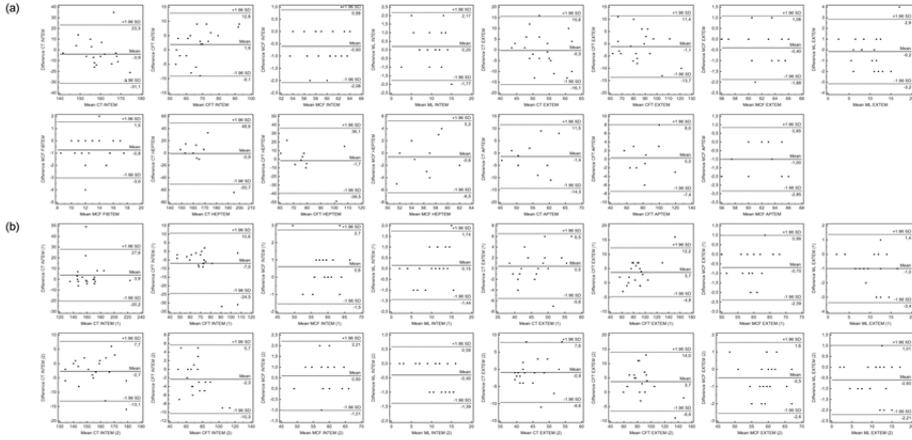
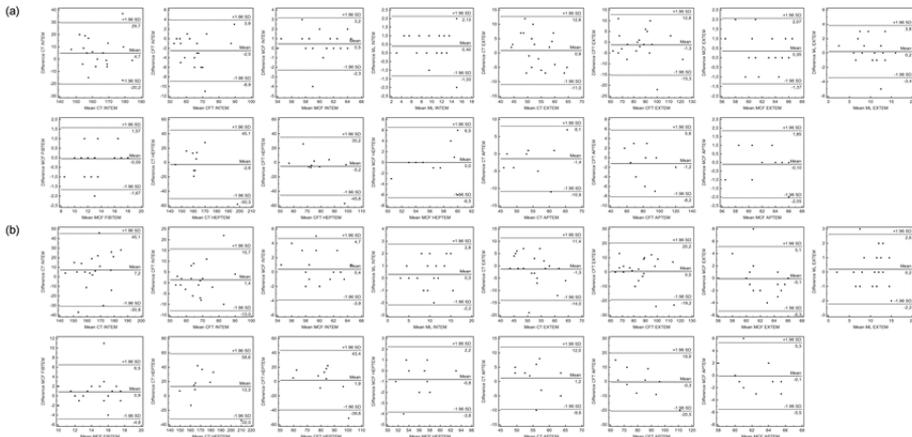


Figure S2. Bland-Altman difference plots for comparisons between (a) morning and afternoon measurements (circadian variation), and (b) if measured four weeks apart. The differences between the two measurements are plotted against the average (mean) of the two measurements (morning measurement, afternoon measurement; day 0, day 20 or time 0, time 1 hour respectively). Horizontal lines are drawn at the mean differences and the limits of agreement (± 1.96 standard deviation).



Discussion

This systematic investigation reveals large differences in the results of some thromboelastometry parameters analyses and lack of homogeneity. Differences appear not only between analysers, but also between the different channels of the same analyser, between morning and afternoon measurements and when four weeks apart measured. Furthermore, there is an inconsistency within individual tests (INTEM, EXTEM, FIBTEM, APTEM, HEPTEM). Homogeneity of measurements for MCF, ML, and LI30 were higher than average.

This is the first study exploring the consistency of thromboelastometry parameters systematically in a joint assessment. We are aware of some preliminary studies addressing similar questions and showing results, which are partially contradictory and partially in line with ours. For example, Chitlur and colleagues investigated the reproducibility between different analysers in platelet rich plasma, pooled plasma samples and factor VIII deficient samples, which were analysed with 11 ROTEM[®]'s in nine laboratories [26]. A coefficient of variation was calculated to be as high as 39-51% for CFT measurements, 13-14% for CT, 15-16% for MCF and 3-16% for α . UK NEQAS (United Kingdom National External Quality Assessment Service) undertook four surveys in which lyophilized plasma samples were sent to 10 centres [23]. A coefficient of variation was calculated as 7-29% in case of MCF and 8-29% in case of CT. In a study by Dick et al. lyophilised plasma samples were sent to 51 centres and a coefficient of variation was calculated as 10-16% for CT, 7% for A20 and 2 – 5% for α [4]. Theusinger et al compared measurements of two analysers in 10 volunteers using analysis of variance (ANOVA). A statistically significant difference was found in some parameters (MCF [INTEM, FIBTEM], CT, CFT and α [INTEM]), but not in others [29]. In another investigation “repeatability” was observed by measuring thromboelastometry serially in 7 citrated whole blood samples of two volunteers one after another [27]. Results were rated as “poor” (with the exception of MCF and α). In contrast, Sørensen reported a coefficient of variation (CV) of 3% regarding MCF, 5-13% (CT) and 7-17% (CFT) [28]. However, information on material and type of determination is lacking. Another study observed reproducibility by “duplicate measurements” of 10 samples on different channels. CV was calculated as < 6% in most parameters and < 15% in CT parameters [29]. Lang et al. observed intra-assay variability by repeating 7 citrated whole blood samples 4 to 15 times on the same device and found a CV between 2 and 13% [30]. Another investigation focussed on monitoring of rFVIIa in haemophilia patients revealed large intra-and inter-patient variation and concluded that the method is unsuitable in this situation [33]. In summary, the results in the mentioned investigations are variable and key questions on reproducibility remained unanswered. The strength of our study is that we used a design that takes several distinct features of the thromboelastometry analysis into account: the vulnerable characteristics of whole blood samples (including thrombocytes, coagulation factors and inhibitors), the multi-channel composition of analysers and the long duration time of measurements. In contrast to previous investigations, we used a systematic approach to avoid possible confounding variables. In addition, we used citrated whole blood samples and not a surrogate material such as plasma samples. A limitation of our study is that it included only healthy volunteers rather than a mixed population of healthy and ill subjects.

Our results provide essential prerequisites for studies on validity of thromboelastometry analysis. Discrimination can be made between parameters and situations, in which thromboelastometry analysis is reproducible and others, in which it is not. This is

an important finding for physicians using the ROTEM® in clinical practice and scientists designing clinical studies alike. Our results may help them to concentrate on the right parameters among the huge number of parameters each measurement yields. This may bring the discussion about clinical value of thromboelastometry analysis forward. Although the various parameters and tests of thromboelastometry analysis represent different aspects of clot formation, every parameter must prove its technical performance (including reproducibility), diagnostic performance, diagnostic impact, organisational impact, and cost effectiveness to qualify for clinical decision making [24]. Clinical studies would be much more informative if they would concentrate on a few, clearly characterised parameters. If confirmed in patient samples as well, MCF and ML are promising parameters to focus on.

In conclusion, systematic investigation reveals large differences in the results of some thromboelastometry parameters and lack of consistency. Clinicians and scientists should be aware of these inconsistencies, look at ROTEM® results mindfully and perhaps focus on parameters with a higher homogeneity such as MCF.

References

1. Haas T (2010) Point of care diagnostic: thromboelastometry (ROTEM(R)). *Wien Klin Wochenschr* 122 Suppl 5: S19-20.
2. Luddington RJ (2005) Thrombelastography/thromboelastometry. *Clin Lab Haematol* 27: 81-90.
3. Fenger-Eriksen C, Jensen TM, Kristensen BS, Jensen KM, Tonnesen E, et al. (2009) Fibrinogen substitution improves whole blood clot firmness after dilution with hydroxyethyl starch in bleeding patients undergoing radical cystectomy: a randomized, placebo-controlled clinical trial. *J Thromb Haemost* 7: 795-802.
4. Dick A, Schwaiger M, Jambor C (2010) [Thromboelastography/-metry and external quality control. Results of a pilot study]. *Hamostaseologie* 30: 91-95.
5. Rugeri L, Levrat A, David JS, Delecroix E, Floccard B, et al. (2007) Diagnosis of early coagulation abnormalities in trauma patients by rotation thrombelastography. *J Thromb Haemost* 5: 289-295.
6. Shore-Lesserson L, Manspeizer HE, DePerio M, Francis S, Vela-Cantos F, et al. (1999) Thromboelastography-guided transfusion algorithm reduces transfusions in complex cardiac surgery. *Anesth Analg* 88: 312-319.
7. Girdauskas E, Kempfert J, Kuntze T, Borger MA, Enders J, et al. (2010) Thromboelastometrically guided transfusion protocol during aortic surgery with circulatory arrest: a prospective, randomized trial. *J Thorac Cardiovasc Surg* 140: 1117-1124 e1112.
8. Fenger-Eriksen C, Tonnesen E, Ingerslev J, Sorensen B (2009) Mechanisms of hydroxyethyl starch-induced dilutional coagulopathy. *J Thromb Haemost* 7: 1099-1105.
9. Tauber H, Innerhofer P, Breikopf R, Westermann I, Beer R, et al. (2011) Prevalence and impact of abnormal ROTEM(R) assays in severe blunt trauma: results of the 'Diagnosis and Treatment of Trauma-Induced Coagulopathy (DIA-TRE-TIC) study'. *Br J Anaesth* 107: 378-387.
10. Weber CF, Gorlinger K, Meininger D, Herrmann E, Bingold T, et al. (2012) Point-of-care testing: a prospective, randomized clinical trial of efficacy in coagulopathic cardiac surgery patients. *Anesthesiology* 117: 531-547.
11. Davenport R, Manson J, De'Ath H, Platton S, Coates A, et al. (2011) Functional definition and characterization of acute traumatic coagulopathy. *Crit Care Med* 39: 2652-2658.
12. Rourke C, Curry N, Khan S, Taylor R, Raza I, et al. (2012) Fibrinogen levels during trauma hemorrhage, response to replacement therapy, and association with patient outcomes. *J Thromb Haemost* 10: 1342-1351.
13. Schochl H, Cotton B, Inaba K, Nienaber U, Fischer H, et al. (2011) FIBTEM provides early prediction of massive transfusion in trauma. *Crit Care* 15: R265.
14. Kozek-Langenecker SA, Afshari A, Albaladejo P, Santullano CA, De Robertis E, et al. (2013) Management of severe perioperative bleeding: Guidelines from the European Society of Anaesthesiology. *Eur J Anaesthesiol* 30: 270-382.
15. Spahn DR, Bouillon B, Cerny V, Coats TJ, Duranteau J, et al. (2013) Management of bleeding and coagulopathy following major trauma: an updated European guideline. *Crit Care* 17: R76.
16. Afshari A, Wikkelsø A, Brok J, Møller AM, Wetterslev J (2011) Thrombelastography (TEG) or thromboelastometry (ROTEM) to monitor haemotherapy versus usual care in patients with massive transfusion. *Cochrane Database Syst Rev*: CD007871.
17. Coakley M, Hall JE, Evans C, Duff E, Billing V, et al. (2011) Assessment of thrombin generation measured before and after cardiopulmonary bypass surgery and its association with postoperative bleeding. *J Thromb Haemost* 9: 282-292.
18. Lee GC, Kicza AM, Liu KY, Nyman CB, Kaufman RM, et al. (2012) Does rotational thromboelastometry (ROTEM) improve prediction of bleeding after cardiac surgery? *Anesthesia and analgesia* 115: 499-506.

19. Wikkelse AJ, Afshari A, Wetterslev J, Brok J, Moeller AM (2011) Monitoring patients at risk of massive transfusion with Thrombelastography or Thromboelastometry: a systematic review. *Acta anaesthesiologica Scandinavica* 55: 1174-1189.
20. Raza I, Davenport R, Rourke C, Platton S, Manson J, et al. (2012) The incidence and magnitude of fibrinolytic activation in trauma patients. *J Thromb Haemost*.
21. Callum JL, Rizoli S (2012) Plasma transfusion for patients with severe hemorrhage: what is the evidence? *Transfusion* 52 Suppl 1: 30S-37S.
22. Chitlur M, Lusher J (2010) Standardization of thromboelastography: values and challenges. *Semin Thromb Hemost* 36: 707-711.
23. Kitchen DP, Kitchen S, Jennings I, Woods T, Walker I (2010) Quality assurance and quality control of thrombelastography and rotational Thromboelastometry: the UK NEQAS for blood coagulation experience. *Semin Thromb Hemost* 36: 757-763.
24. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 46: 1041-1050.
25. Bruns DE, Huth EJ, Magid E, Young DS (2000) Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. *Clin Chem* 46: 893-895.
26. Chitlur M, Sorensen B, Rivard GE, Young G, Ingerslev J, et al. (2011) Standardization of thromboelastography: a report from the TEG-ROTEM working group. *Haemophilia* 17: 532-537.
27. Venema LF, Post WJ, Hendriks HG, Huet RC, de Wolf JT, et al. (2010) An assessment of clinical interchangeability of TEG and RoTEM thromboelastographic variables in cardiac surgical patients. *Anesth Analg* 111: 339-344.
28. Sorensen B, Johansen P, Christiansen K, Woelke M, Ingerslev J (2003) Whole blood coagulation thrombelastographic profiles employing minimal tissue factor activation. *J Thromb Haemost* 1: 551-558.
29. Theusinger OM, Nurnberg J, Asmis LM, Seifert B, Spahn DR (2010) Rotation thromboelastometry (ROTEM) stability and reproducibility over time. *Eur J Cardiothorac Surg* 37: 677-683.
30. Lang T, Bauters A, Braun SL, Potzsch B, von Pape KW, et al. (2005) Multi-centre investigation on reference ranges for ROTEM thromboelastometry. *Blood Coagul Fibrinolysis* 16: 301-310.
31. Lang T, von Depka M (2006) [Possibilities and limitations of thrombelastometry/-graphy]. *Hamostaseologie* 26: S20-29.
32. Dargaud Y, Sorensen B, Shima M, Hayward C, Srivastava A, et al. (2012) Global haemostasis and point of care testing. *Haemophilia* 18 Suppl 4: 81-88.
33. Kenet G, Stenmo CB, Blemings A, Wegert W, Goudemand J, et al. (2010) Intra-patient variability of thromboelastographic parameters following in vivo and ex vivo administration of recombinant activated factor VII in haemophilia patients. A multi-centre, randomised trial. *Thromb Haemost* 103: 351-359.

CHAPTER 4

Impact of changes in haematocrit level and platelet count on thromboelastometry parameters

Michael Nagler, Silvio Kathriner, Lucas M. Bachmann, Walter A. Wuillemin
Thrombosis Research 2013, 131: 249–253

Abstract

Introduction: To what extent haematocrit levels (Hct) and platelet counts (PLT) influence the measurement of parameters of thromboelastometry when assessed with the ROTEM® device is unclear. We investigated to what extent thromboelastometry measurements depend on Hct and PLT.

Materials and Methods: Whole blood samples were taken for in-vitro preparations of mixtures with three different levels of PLT and a varying Hct. Maximum clot firmness (MCF), clotting time (CT), clot formation time (CFT) and alpha angle (α) for INTEM, EXTEM, FIBTEM and APTEM was recorded.

Results: Measurements depended substantially on Hct and PLT. MCF readings were systematically lower with increasing Hct (0.2 vs. 0.4: -7.8 (-8.3 to -7.2); $p < 0.001$, 0.2 vs. 0.55: -14.5 (-17.3 to -14.3); $p < 0.001$) but higher with increasing PLT (50 vs. $125 \times 10^9/l$: 8.2 (4.2 to 12.3); $p = 0.005$, 50 vs. $250 \times 10^9/l$: 12.0 (7.2 to 16.8); $p = 0.002$). CT readings were systematically higher with increasing Hct (0.2 vs. 0.4: 9.2 (6.2 to 12.1); $p = 0.001$, 0.2 vs. 0.55: 38.2 (21.5 to 54.9); $p = 0.003$) while increasing PLT had no influence. CFT readings were also systematically higher with increasing Hct (0.2 vs. 0.4: 83.8 (40.2 to 127.6); $p = 0.006$, 0.2 vs. 0.55: 226.2 (110.7 to 341.7); $p = 0.006$) but systematically lower with increasing PLT (50 vs. $125 \times 10^9/l$: -144.0 (-272.3 to -15.6); $p = 0.036$, 50 vs. $250 \times 10^9/l$: -189.2 (-330.4 to -48.0); $p = 0.02$); readings of the alpha angle showed a similar pattern.

Conclusions: Our results suggest that readings of thromboelastometry parameters need to be adjusted by Hct and PLT to avoid potential confounding and misinterpretations in clinical practice.

Introduction

Thromboelastometry analysis (ROTEM®) is a point-of-care viscoelastic test comparable to thromboelastography (TEG®), which is widely used in acute clinical care settings. Operators are attracted by the ease of use and the fast production of graphs that are easy to grasp, displaying the results [1,2]. Due to its claim to provide global information on all parts of haemostasis including platelet function, coagulation cascade, cross-linking of fibrin and fibrinolysis, it is seen as an innovation in the setting of acute bleeding [3,4] and several other clinical situations [5-10]. However, while thromboelastometry has been thoroughly evaluated in perioperative and trauma settings for the detection of coagulopathies, to guide substitution therapy and to reduce unnecessary transfusions, key factors influencing the measurement of thromboelastometry parameters are still incompletely understood [3,11-14].

For example, the impact of important factors of influence, such as haematocrit level (Hct) and platelet count (PLT), on thromboelastometry parameters as assessed with the ROTEM® device have not been studied in detail. Investigations in different clinical and experimental settings found associations between Hct and PLT with some thromboelastometry parameters [11,15-22]. Other investigations indicate that changing Hct affects thromboelastometry parameters when used as trigger for fibrinogen supplementation [15]. However, due to design and the clinical context of the studies, confounding variables such as dilution phenomena have possibly affected the results. Nevertheless, the type and the extent of changes on particular thromboelastometry parameters due to isolated, specific changes of Hct and PLT (alone and in combination) are uncertain. The problem of Hct and PLT as possible confounders is particularly important in the context of massive bleeding, where rapid changes of Hct and PLT are common [23,24]. Therefore, changes of Hct and PLT may have contributed to the uncertainty of the clinical value, which has been questioned by some authors and two systematic reviews [23,25-28]. Sound evidence on factors of influence of a laboratory test are essential pre-requisite for clinical studies and may facilitate large studies, which are valid and are able to answer questions on clinical value [29]. Therefore, the aim of the present in-vitro evaluation study was to investigate the impact of changes in Hct and PLT on a broad spectrum of thromboelastometry parameters systematically in a strictly experimental design excluding additional confounding variables such as dilution phenomena

Materials and methods

Study design

Citrated whole blood samples were taken from five healthy volunteers for in-vitro preparations (45 measurements). Individuals did not have any known disease, were taking no medication with an effect on haemostatic parameters and had no history of bleeding or thromboembolic events. As an inclusion criterion, baseline measurements of haematocrit, PLT, white blood count, routine haemostatic parameters including fibrinogen level and thromboelastometry parameters were within reference range. The study was approved by the local ethical review board (Kantonale Ethikkommission Luzern) and all participants provided written informed consent.

Sampling

Blood withdrawal was done from cubital vein under standardised conditions. A 21-gauge butterfly cannula and plastic syringes (Monovette[®], Sarstedt, Nümbrecht, Germany) containing 1 ml trisodium citrate (0.106 mol/l) for 9 ml of blood were used. Citrated sample for thromboelastometry was collected after EDTA sample (for determining platelet count) to avoid tissue factor contamination. Tourniquet application was limited to a maximum of one minute and removed before blood withdrawal to avoid platelet activation. Samples were processed immediately by hand to preparation. Samples for baseline measurements were collected in advance.

In-vitro preparations

For each blood withdrawal, platelet rich plasma (PRP), platelet poor plasma (PPP) and mixtures with three different levels of PLT (50, 125 or 250x10⁹/l respectively) and a varying haematocrit (0.2, 0.4 and 0.55) were prepared. Preparations were done in analogy to current standards of light transmission aggregometry [30]. PRP was made by centrifugation at 150 x g for 15 minutes, PPP by centrifugation at 1500 x g for 15 minutes. Blood samples with an intended PLT of 50x10⁹/l were centrifuged at 150 x g for 15 minutes, plasma fraction removed and PPP added to the anticipated Hct. Samples with an intended PLT of 125 or 250x10⁹/l were centrifuged at 1500 x g for 15 minutes, plasma fraction removed and PPP and/or PRP added to the anticipated haematocrit and PLT level. Proportions of PPP and/or PRP necessary were determined in a pilot study. Samples were resuspended by gentle agitation. If Hct and/or PLT deviated substantially from intended levels, a limited number of correction measures (ad maximum 3 times) by means of adding PPP or PRP respectively were allowed. Fibrinogen level (as measured by Clauss method) was held constant, as no dilution of the plasma has been done. Each sample was handled identically to limit potential influences such as platelet activation. Preparation was done in the original syringe (only PRP and PPP was transferred into separate tubes), temperature was held stable at 20-24°C, no brake

were used during centrifugation and manipulations were limited to a minimal amount. Samples have been continually agitated and processed immediately to centrifugation, preparation and analysis.

Analysis

PLT, Hct and white blood count were determined using XE 5000 (Sysmex AG, Horgen, Switzerland). Thromboelastometry was conducted using three ROTEM[®] analysers (ROTEM[®] delta; Pentapharm GmbH, Munich, Germany). Preparation of blood samples and thromboelastometry procedures were performed according to the manufacturer's instructions by following an automated electronic pipette program. We analysed the following parameters in INTEM, EXTEM and APTTEM test: clotting time (CT, s), clot formation time (CFT, s), maximum clot firmness (MCF, mm) and the alpha angle (α). In FIBTEM test we analysed maximum clot firmness (MCF, mm). Details of the ROTEM[®] analysis, parameters and activating agents have been described in detail elsewhere [11,31]. Time between in-vitro preparations and analysis of the samples was strictly held at 10-15 minutes.

Statistical analysis

For each of the four parameters CT, CFT, MCF and alpha angle we fitted a mixed linear model entering PLT and Hct using two indicator variates. For the CT, CFT and alpha angle model we entered two indicator variates for the INTEM, EXTEM, and APTTEM test. For the MCF assessment we also entered the FIBTEM test using a third indicator variate. For all analyses, the subject variable was entered as a random factor to take into account that the same blood sample was used for various measurements. All analyses were performed using the Stata 11.1 statistics software package (StataCorp LP, College Station, TX, USA).

Results

Hct and PLT realised with the different samples are displayed in **Table 1**. Thromboelastometry values depended substantially on Hct and PLT; the distribution of the measurements with regard to the individual tests (INTEM, EXTEM, FIBTEM, APTTEM) are displayed in **Table 2 to Table 5**.

MCF readings were systematically lower with increasing Hct (0.2 vs. 0.4: -7.8 (-8.3 to -7.2); $p < 0.001$, 0.2 vs. 0.55: -14.5 (-17.3 to -14.3); $p < 0.001$) but systematically higher with increasing PLT (50 vs. $125 \times 10^9/l$: 8.2 (4.2 to 12.3); $p = 0.005$, 50 vs. $250 \times 10^9/l$: 12.0 (7.2 to 16.8); $p = 0.002$).

CT readings were systematically higher with increasing Hct (0.2 vs. 0.4: 9.2 (6.2 to 12.1); $p = 0.001$, 0.2 vs. 0.55: 38.2 (21.5 to 54.9); $p = 0.003$) while increasing PLT had no

influence (50 vs. $125 \times 10^9/l$: -4.7 (-18.4 to 8.9); $p=0.392$, 50 vs. $250 \times 10^9/l$: -4.2 (-18.3 to 9.9); $p<0.001$).

CFT readings were also systematically higher with increasing Hct (0.2 vs. 0.4: 83.8 (40.2 to 127.6); $p=0.006$, 0.2 vs. 0.55: 226.2 (110.7 to 341.7); $p=0.006$) but systematically lower with increasing PLT (50 vs. $125 \times 10^9/l$: -144.0 (-272.3 to -15.6); $p=0.036$, 50 vs. $250 \times 10^9/l$: -189.2 (-330.4 to -48.0); $p=0.02$).

Readings of the alpha angle showed a similar pattern as the MCF readings. (For Hct increases: 0.2 vs. 0.4: -13.3 (-17.9 to -8.7); $p=0.001$, 0.2 vs. 0.55: -26.0 (-32.5 to -19.5); $p<0.001$. For increasing PLT: (50 vs. $125 \times 10^9/l$: 7.1 (1.0 to 13.2); $p=0.032$, 50 vs. $250 \times 10^9/l$: 10.8 (4.4 to 17.2); $p=0.009$).

Table 1: Haematocrit and platelet count realised in the sample series with regard to the intended levels

| Parameter | Intended level | Level realised median (range) |
|------------------------------------|----------------|-------------------------------|
| Haematocrit | 0.2 | 0.20 (0.17-0.22)* |
| | 0.4 | 0.39 (0.36-0.41)* |
| | 0.55 | 0.53 (0.50-0.57)* |
| Platelet count ($\times 10^9/l$) | 50 | 58 (34-81)* |
| | 125 | 150 (95-186)* |
| | 250 | 242 (201-282)* |

* n=15

Table 2: Effects of increasing haematocrit level on MCF at different platelet counts* and effects of increasing platelet count on MCF at different haematocrit levels#

| Test | Haematocrit level | MCF | | |
|--------|-------------------|---------------------------|--------------------------|--------------------------|
| | | <i>mm; median (range)</i> | | |
| | | <i>50 G platelets/l</i> | <i>125 G platelets/l</i> | <i>250 G platelets/l</i> |
| INTEM | 0.2 | 46 (42-49) | 59 (55-61) | 65 (62-68) |
| | 0.4 | 43 (35-46) | 51 (51-54) | 56 (54-59) |
| | 0.55 | 31 (29-40) | 45 (42-51) | 50 (44-56) |
| EXTEM | 0.2 | 45 (44-52) | 61(59-62) | 65 (61-66) |
| | 0.4 | 39 (32-49) | 49 (46-54) | 54 (53-57) |
| | 0.55 | 34 (29-39) | 40 (36-49) | 45 (41-51) |
| FIBTEM | 0.2 | 16 (14-22) | 18 (12-21) | 18 (17-19) |
| | 0.4 | 10 (8-15) | 11 (10-12) | 11 (11-12) |
| | 0.55 | 7 (5-9) | 9 (7-11) | 9 (8-11) |
| APTEM | 0.2 | 44 (40-52) | 59 (55-62) | 65 (59-65) |
| | 0.4 | 38 (31-48) | 45 (45-51) | 53 (47-60) |
| | 0.55 | 33 (24-34) | 38 (33-43) | 44 (34-53) |

* displayed in columns of the particular tests; # displayed in rows of the particular tests

Table 3: Effects of increasing haematocrit level on CT at different platelet counts* and effects of increasing platelet count on CT at different haematocrit levels#

| Test | Haematocrit level | CT | | |
|-------|-------------------|-------------------|-------------------|-------------------|
| | | s; median (range) | | |
| | | 50 G platelets/l | 125 G platelets/l | 250 G platelets/l |
| INTEM | 0.2 | 169 (153–178) | 163 (132–179) | 162 (139–183) |
| | 0.4 | 170 (151–188) | 169 (134–187) | 165 (148–179) |
| | 0.55 | 203 (171–266) | 195 (184–236) | 198 (175–239) |
| EXTEM | 0.2 | 45 (41–49) | 36 (35–43) | 46 (36–51) |
| | 0.4 | 52 (41–106) | 57 (47–65) | 56 (39–76) |
| | 0.55 | 69 (46–86) | 86 (43–130) | 68 (55–115) |
| APTEM | 0.2 | 48 (38–52) | 39 (37–48) | 43 (40–53) |
| | 0.4 | 54 (41–60) | 47 (26–64) | 50 (33–62) |
| | 0.55 | 75 (50–170) | 68 (51–130) | 78 (55–129) |

* displayed in columns of the particular tests; # displayed in rows of the particular tests

Table 4: Effects of increasing haematocrit level on CFT at different platelet counts* and effects of increasing platelet count on CT at different haematocrit levels#

| Test | Haematocrit level | CFT | | |
|-------|-------------------|-------------------|-------------------|-------------------|
| | | s; median (range) | | |
| | | 50 G platelets/l | 125 G platelets/l | 250 G platelets/l |
| INTEM | 0.2 | 129 (99–166) | 66 (54–72) | 46 (40–54) |
| | 0.4 | 200 (128–335) | 99 (87–123) | 79 (72–110) |
| | 0.55 | 460 (201–647) | 170 (117–223) | 125 (93– 238) |
| EXTEM | 0.2 | 158 (151–202) | 85 (70–89) | 77 (61–78) |
| | 0.4 | 255 (181–502) | 161 (113–204) | 130 (98–159) |
| | 0.55 | 452 (299–749) | 240 (174–470) | 199 (153– 271) |
| APTEM | 0.2 | 186 (135–244) | 93 (83–110) | 70 (69–94) |
| | 0.4 | 277 (194–488) | 198 (135–231) | 123 (98–245) |
| | 0.55 | 449 (363–1174) | 352 (241–657) | 208 (144– 344) |

* displayed in columns of the particular tests; # displayed in rows of the particular tests

Table 5: Effects of increasing haematocrit level on alpha angle (α) at different platelet counts* and effects of increasing platelet count on CT at different haematocrit levels#

| Test | Haematocrit level | alpha angle (α) degree; median (range) | | |
|-------|-------------------|---|-------------------|-------------------|
| | | 50 G platelets/l | 125 G platelets/l | 250 G platelets/l |
| INTEM | 0.2 | 75 (74–78) | 78 (77–80) | 80 (79–82) |
| | 0.4 | 59 (50–73) | 71 (66–75) | 75 (70–76) |
| | 0.55 | 40 (31–57) | 60 (52–67) | 66 (53–71) |
| EXTEM | 0.2 | 74 (69–76) | 79 (75–81) | 74 (74–79) |
| | 0.4 | 52 (31–67) | 60 (53–68) | 65 (61–73) |
| | 0.55 | 37 (26–61) | 55 (32–58) | 55 (46–61) |
| APTEM | 0.2 | 72 (67–79) | 75 (71–78) | 76 (74–76) |
| | 0.4 | 50 (47–67) | 60 (50–73) | 66 (48–81) |
| | 0.55 | 37 (32–65) | 46 (27–55) | 57 (46–65) |

* displayed in columns of the particular tests; # displayed in rows of the particular tests

Discussion

Isolated changes of Hct and PLT have a substantial impact on a broad range of thromboelastometry parameters. Our results show that Hct and PLT are major, independent factors of influence of thromboelastometry parameters. Thus, this report complies with the call for critical evaluations needed for laboratory tests used in clinical practice.

This study corroborates previous investigations, which found associations in different clinical and experimental settings. A prolongation of clotting times (CT, CFT, α) and impairment of MCF (INTEM, EXTEM) was seen in an equine animal model of transient polycythaemia (n=6) [20]. In another animal study of 127 healthy dogs moderate correlations were found between some thromboelastometry parameters and PLT (rank correlation coefficient r_s : 0.16 – 0.55) as well as Hct (r_s : 0.17 – 0.60) [16]. In a study of 90 trauma patients a moderate correlation was found between some thromboelastometry parameters and PLT (CFT-EXTEM: $r=0.33$; A15-EXTEM: $r=0.56$; CFT-INTEM: $r=0.32$; CFT-INTEM: $r=0.57$) [11]. In another case-control study, thromboelastometry profiles of 40 patients with sideropenic anaemia were compared with healthy volunteers and a statistically significant impairment of MCF (INTEM, EXTEM, FIBTEM) was found [22]. In the same study, authors did also in-vitro manipulations by adding PPP to whole blood samples with comparable results to the case-control study. Another in-vitro investigation on plasma (PPP or PRP respectively) of healthy volunteers found an increase of MCF in EXTEM and FIBTEM with escalating PLT [19]. Correlations between PLT and MCF or A10 respectively were also found in investigations of patients undergoing orthotopic liver

transplantation [17,21]. A reduction of CFT (INTEM) and increase of MCF (INTEM) was seen in another investigation of 10 thrombocytopenic haematological patients after platelet transfusion [18]. Other investigations pointed out that changing Hct affects thromboelastometry parameters when used as trigger for fibrinogen supplementation [15]. Associations between thromboelastography parameters (when assessed with the TEG® device) and changes in Hct [32] or PLT [33,34] were similar.

What is the strength of our investigation? In contrast to previous investigations, we studied blood samples of healthy volunteers applying a strictly scientific protocol, which allowed an unbiased and rigorous examination of effects of Hct and PLT on a broad range of thromboelastometry parameters. Most previous studies were done in animals, in particular clinical situations or used non-physiological blood mixtures – in settings that are susceptible to confounding. Our in-vitro setting allowed examining, the isolated effect of Hct and PLT on thromboelastometry parameters in a valid fashion excluding the possibility that dilution phenomena and other circumstances influenced the results. The consistency of effects across all assessed thromboelastometry parameters is another strong point of this paper.

What are the limitations of the study? Arguably, the varying amounts of fibrinogen may have influenced the results. However, no dilution of the plasma was done and the fibrinogen concentrations in a random sample of 5 aliquots (representing 45 measurements) were equal when using the Clauss method. Moreover, our blood samples were from healthy volunteers making it very unlikely that preparations produced critical fibrinogen levels.

Our results emphasise, that Hct and PLT may act as relevant confounding variables when thromboelastometry is used in clinical practice and scientific investigations. Two systematic reviews and several papers raised concerns about the clinical value of thromboelastometry analysis in the setting of acute bleeding [25-28]. Other investigations indicate that changing Hct levels affect thromboelastometry when used as a trigger for fibrinogen substitution [15]. Furthermore, changes of Hct and PLT in the range investigated are not necessarily associated with a change in bleeding tendency. In contrast to the ‘hypocoagulable pattern’ we have found, there are various investigations indicating the prothrombotic potential of red blood cells [35]. Finally, whether or not the effects shown in this paper actually have an impact on clinical bleeding tendency needs to be addressed in properly designed clinical trials. We believe, until further investigations are available, alterations of thromboelastometry parameters due to changes of Hct and PLT should be interpreted primarily as a consequence of the laboratory method used. Sound evidence about parameters influencing a laboratory test is an essential pre-requisite for large clinical studies and must be considered while designing them [29]. Bearing in mind that the effects on thromboelastometry parameters were

similar to those observed in clinical studies [23,36-39] we think that future investigations should take Hct and PLT as possible confounding variables into account.

What are the implications for clinical practice? We believe that thromboelastometry analysis (using the ROTEM[®] device) has a role to play in the management of patients at risk for massive bleeding, particularly regarding the identification of coagulopathy due to hyperfibrinolysis and critically low fibrinogen levels. Up to know, however, solid evidence identifying the areas of optimal use remains sparse. In our view, thromboelastometry analysis should be limited to screening situations, while therapy control should be discouraged until new evidence regarding the problems identified in this paper becomes available. Furthermore, we think that changes of thromboelastometry parameters need to be interpreted with caution and ideally in the context of changing Hct and PLT.

In conclusion, our results suggest that readings of thromboelastometry parameters need to be adjusted by Hct and PLT to avoid potential confounding and misinterpretations in clinical practice and future investigations.

References

1. Haas T (2010) Point of care diagnostic: thromboelastometry (ROTEM(R)). *Wien Klin Wochenschr* 122 Suppl 5: S19-20.
2. Luddington RJ (2005) Thrombelastography/thromboelastometry. *Clin Lab Haematol* 27: 81-90.
3. Fenger-Eriksen C, Jensen TM, Kristensen BS, Jensen KM, Tonnesen E, et al. (2009) Fibrinogen substitution improves whole blood clot firmness after dilution with hydroxyethyl starch in bleeding patients undergoing radical cystectomy: a randomized, placebo-controlled clinical trial. *J Thromb Haemost* 7: 795-802.
4. Dick A, Schwaiger M, Jambor C (2010) [Thromboelastography/-metry and external quality control. Results of a pilot study]. *Hamostaseologie* 30: 91-95.
5. Huissoud C, Carrabin N, Benchaib M, Fontaine O, Levrat A, et al. (2009) Coagulation assessment by rotation thrombelastometry in normal pregnancy. *Thromb Haemost* 101: 755-761.
6. Kenet G, Stenmo CB, Blemings A, Wegert W, Goudemand J, et al. (2010) Intra-patient variability of thromboelastographic parameters following in vivo and ex vivo administration of recombinant activated factor VII in haemophilia patients. A multi-centre, randomised trial. *Thromb Haemost* 103: 351-359.
7. Carroll RC, Craft RM, Whitaker GL, Snider CC, Kirby RK, et al. (2007) Thrombelastography monitoring of resistance to enoxaparin anticoagulation in thrombophilic pregnancy patients. *Thromb Res* 120: 367-370.
8. Sorensen B, Asvaldsdottir HS, Gudmundsdottir BR, Onundarson PT (2009) The combination of recombinant factor VIIa and fibrinogen correct clotting ex vivo in patient samples obtained following cardiopulmonary bypass surgery. *Thromb Res* 124: 695-700.
9. Tripodi A, Primignani M, Chantarangkul V, Viscardi Y, Dell'Era A, et al. (2009) The coagulopathy of cirrhosis assessed by thromboelastometry and its correlation with conventional coagulation parameters. *Thromb Res* 124: 132-136.
10. van Veen JJ, Gatt A, Bowyer AE, Cooper PC, Kitchen S, et al. (2009) Calibrated automated thrombin generation and modified thromboelastometry in haemophilia A. *Thromb Res* 123: 895-901.
11. Rugeri L, Levrat A, David JS, Delecroix E, Floccard B, et al. (2007) Diagnosis of early coagulation abnormalities in trauma patients by rotation thrombelastography. *J Thromb Haemost* 5: 289-295.
12. Shore-Lesserson L, Manspeizer HE, DePerio M, Francis S, Vela-Cantos F, et al. (1999) Thromboelastography-guided transfusion algorithm reduces transfusions in complex cardiac surgery. *Anesth Analg* 88: 312-319.
13. Girdauskas E, Kempfert J, Kuntze T, Borger MA, Enders J, et al. (2010) Thromboelastometrically guided transfusion protocol during aortic surgery with circulatory arrest: a prospective, randomized trial. *J Thorac Cardiovasc Surg* 140: 1117-1124 e1112.
14. Fenger-Eriksen C, Tonnesen E, Ingerslev J, Sorensen B (2009) Mechanisms of hydroxyethyl starch-induced dilutional coagulopathy. *J Thromb Haemost* 7: 1099-1105.
15. Ogawa S, Szlam F, Bolliger D, Nishimura T, Chen EP, et al. (2012) The Impact of Hematocrit on Fibrin Clot Formation Assessed by Rotational Thromboelastometry. *Anesthesia and analgesia*.
16. Smith SA, McMichael MA, Gilor S, Galligan AJ, Hoh CM (2012) Correlation of hematocrit, platelet concentration, and plasma coagulation factors with results of thromboelastometry in canine whole blood samples. *American journal of veterinary research* 73: 789-798.
17. Herbstreit F, Winter EM, Peters J, Hartmann M (2010) Monitoring of haemostasis in liver transplantation: comparison of laboratory based and point of care tests. *Anaesthesia* 65: 44-49.
18. Flisberg P, Rundgren M, Engstrom M (2009) The effects of platelet transfusions evaluated using rotational thromboelastometry. *Anesth Analg* 108: 1430-1432.
19. Lang T, Johanning K, Metzler H, Piepenbrock S, Solomon C, et al. (2009) The effects of fibrinogen levels on thromboelastometric variables in the presence of thrombocytopenia. *Anesth Analg* 108: 751-758.

20. McMichael M, Smith SA, McConachie EL, Lascola K, Wilkins PA (2011) In-vitro hypocoagulability on whole blood thromboelastometry associated with in-vivo expansion of red cell mass in an equine model. *Blood coagulation & fibrinolysis: an international journal in haemostasis and thrombosis* 22: 424-430.
21. Rouillet S, Pillot J, Freyburger G, Blais M, Quinart A, et al. (2010) Rotation thromboelastometry detects thrombocytopenia and hypofibrinogenaemia during orthotopic liver transplantation. *Br J Anaesth* 104: 422-428.
22. Spiezia L, Radu C, Marchioro P, Bertini D, Rossetto V, et al. (2008) Peculiar whole blood rotation thromboelastometry (Rotem) profile in 40 sideropenic anaemia patients. *Thromb Haemost* 100: 1106-1110.
23. Coakley M, Hall JE, Evans C, Duff E, Billing V, et al. (2011) Assessment of thrombin generation measured before and after cardiopulmonary bypass surgery and its association with postoperative bleeding. *J Thromb Haemost* 9: 282-292.
24. Frith D, Goslings JC, Gaarder C, Maegele M, Cohen MJ, et al. (2010) Definition and drivers of acute traumatic coagulopathy: clinical and experimental investigations. *Journal of thrombosis and haemostasis: JTH* 8: 1919-1925.
25. Wikkelsoe AJ, Afshari A, Wetterslev J, Brok J, Moeller AM (2011) Monitoring patients at risk of massive transfusion with Thrombelastography or Thromboelastometry: a systematic review. *Acta anaesthesiologica Scandinavica* 55: 1174-1189.
26. Afshari A, Wikkelso A, Brok J, Moller AM, Wetterslev J (2011) Thrombelastography (TEG) or thromboelastometry (ROTEM) to monitor haemotherapy versus usual care in patients with massive transfusion. *Cochrane Database Syst Rev*: CD007871.
27. Lee GC, Kicza AM, Liu KY, Nyman CB, Kaufman RM, et al. (2012) Does rotational thromboelastometry (ROTEM) improve prediction of bleeding after cardiac surgery? *Anesthesia and analgesia* 115: 499-506.
28. Raza I, Davenport R, Rourke C, Platton S, Manson J, et al. (2012) The incidence and magnitude of fibrinolytic activation in trauma patients. *Journal of thrombosis and haemostasis: JTH*.
29. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 46: 1041-1050.
30. Cattaneo M, Hayward CP, Moffat KA, Pugliano MT, Liu Y, et al. (2009) Results of a worldwide survey on the assessment of platelet function by light transmission aggregometry: a report from the platelet physiology subcommittee of the SSC of the ISTH. *Journal of thrombosis and haemostasis: JTH* 7: 1029.
31. Chitlur M, Sorensen B, Rivard GE, Young G, Ingerslev J, et al. (2011) Standardization of thromboelastography: a report from the TEG-ROTEM working group. *Haemophilia* 17: 532-537.
32. Iselin BM, Willimann PF, Seifert B, Casutt M, Bombeli T, et al. (2001) Isolated reduction of haematocrit does not compromise in vitro blood coagulation. *Br J Anaesth* 87: 246-249.
33. Bowbrick VA, Mikhailidis DP, Stansby G (2003) Influence of platelet count and activity on thromboelastography parameters. *Platelets* 14: 219-224.
34. Roeloffzen WW, Kluin-Nelemans HC, Mulder AB, de Wolf JT (2010) Thrombocytopenia affects plasmatic coagulation as measured by thrombelastography. *Blood Coagul Fibrinolysis* 21: 389-397.
35. Horne MK, 3rd, Cullinane AM, Merryman PK, Hoddeson EK (2006) The effect of red blood cells on thrombin generation. *Br J Haematol* 133: 403-408.
36. Davidson SJ, McGrowder D, Roughton M, Kelleher AA (2008) Can ROTEM thromboelastometry predict postoperative bleeding after cardiac surgery? *J Cardiothorac Vasc Anesth* 22: 655-661.
37. Davenport R, Manson J, De'Ath H, Platton S, Coates A, et al. (2011) Functional definition and characterization of acute traumatic coagulopathy. *Crit Care Med* 39: 2652-2658.
38. Schochl H, Cotton B, Inaba K, Nienaber U, Fischer H, et al. (2011) FIBTEM provides early prediction of massive transfusion in trauma. *Crit Care* 15: R265.
39. Tauber H, Innerhofer P, Breitkopf R, Westermann I, Beer R, et al. (2011) Prevalence and impact of abnormal ROTEM(R) assays in severe blunt trauma: results of the 'Diagnosis and Treatment of Trauma-Induced Coagulopathy (DIA-TRE-TIC) study'. *Br J Anaesth* 107: 378-387.

APPENDIX

Thromboelastometry changes in myeloproliferative neoplasms: surrogate for a procoagulant haemostatic imbalance or a consequence of technical reasons?

Comment on: A. Tripodi et. al *Ann Hematol* (2013) 92:1633-1639

Michael Nagler, Hugo ten Cate

Annals of Hematology 2014, 93(10): 1781-2

Dear Editor,

With great interest, we read the article by Tripodi et al [1] investigating the changes of global hemostatic parameters in patients with myeloproliferative neoplasms (MPN). The authors aimed to study whether global tests would be able to detect a “procoagulant imbalance” that correlates with an increased risk for thromboembolism in patients with MPN, and to propose a measure for risk assessment and follow-up in clinical trials. This issue is in-line with efforts to establish global measures of bleeding and thrombotic risks in a variety of clinical situations [2,3]. However, the most important questions to give an answer on this issue are: (1) Do the changes represent the thrombotic risk or are they a consequence of confounding variables? (2) Is the degree of changes relevant? We feel that these issues require critical appraisal.

The authors compared thromboelastometry measurements in patients with essential thrombocythemia (ET), primary myelofibrosis (IMF), and polycythemia vera (PV) with healthy controls. The median platelet count (PLT) in ET and IMF patients was much higher than in PV patients and healthy controls ([1], Table 1). The influence of PLT on thromboelastometry parameters is well established [4] and our group was able to show that PLT is a major factor of influence, independent from other factors that may affect coagulation (**Figure 1**) [5]. Thus, it seems likely that observed changes are the result of a varying PLT and not an “intrinsic procoagulant imbalance”. This is promoted by the fact that no differences have been found in PV patients ([1], Figure 3).

Another issue is that the investigation is based on the assumption, that patients with MPN have a “procoagulant imbalance” *in general*. Most patients were on antiplatelet therapy, phlebotomy and cytoreductive therapy, which are effective measures to reduce the thromboembolic risk in MPN patients [6]. We are convinced that a study aiming to propose a measure for risk assessment and follow-up must be done using clinical outcomes such as thromboembolic events or at least established risk factors of thromboembolic events in MPN patients [7].

The authors report on two out of nine parameters of INTEM test. A median MCF of 68 mm in ET and IMF (range 56 – 76 and 57 – 78 respectively) was reported versus 61 mm in healthy controls (range 51 – 68 mm). CFT was 47 sec (30 – 91) in ET, 53 sec (34 – 104) in IMF and 71 sec in controls (range 47 – 121). The differences were stated to be statistically significant ($p < 0.05$), but it was not reported if the analysis was corrected for multiple testing (nine parameters of INTEM test). Furthermore, given the large variability and recognized problems in reproducibility [8-10], it has to be discussed which degree of changes are considered as relevant to promote its use in clinical practice and scientific inquiry.

Development of global measures of bleeding and thrombotic risks is an important issue in thrombosis and hemostasis research. However, all validity aspects have to be mentioned to be of value in scientific inquiry and clinical practice.

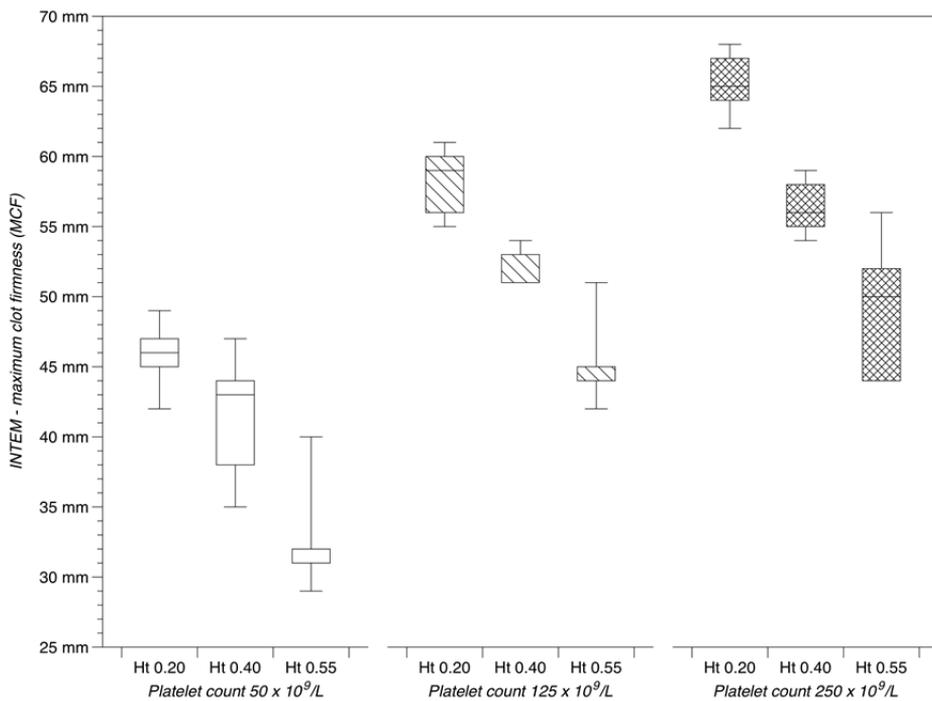


Figure 1: Changes of thromboelastometry parameters with increasing platelet count. Results of an in-vitro investigation studying the impact of hematocrit level (Ht) and platelet count on thromboelastometry parameters [5] are shown (maximum clot firmness [MCF] of the INTEM test with varying platelet counts at different Ht; unpublished figure). Box plots display median, inter-quartile range, and range.

References

1. Tripodi A, Chantarangkul V, Gianniello F, Clerici M, Lemma L, et al. (2013) Global coagulation in myeloproliferative neoplasms. *Ann Hematol* 92: 1633-1639.
2. Dargaud Y, Sorensen B, Shima M, Hayward C, Srivastava A, et al. (2012) Global haemostasis and point of care testing. *Haemophilia* 18 Suppl 4: 81-88.
3. Ten Cate H (2012) Thrombin generation in clinical conditions. *Thromb Res* 129: 367-370.
4. Lier H, Vorweg M, Hanke A, Gorlinger K (2013) Thromboelastometry guided therapy of severe bleeding. Essener Runde algorithm. *Hamostaseologie* 33: 51-61.
5. Nagler M, Kathriner S, Bachmann LM, Wuillemin WA (2013) Impact of changes in haematocrit level and platelet count on thromboelastometry parameters. *Thromb Res* 131: 249-253.
6. Tefferi A (2012) Polycythemia vera and essential thrombocythemia: 2012 update on diagnosis, risk stratification, and management. *Am J Hematol* 87: 285-293.
7. Barbui T, Finazzi G, Carobbio A, Thiele J, Passamonti F, et al. (2012) Development and validation of an International Prognostic Score of thrombosis in World Health Organization-essential thrombocythemia (IPSET-thrombosis). *Blood* 120: 5128-5133; quiz 5252.
8. Nagler M, ten Cate H, Kathriner S, Casutt M, Bachmann LM, et al. (2014) Consistency of thromboelastometry analysis under scrutiny: Results of a systematic evaluation within and between analysers. *Thromb Haemost* (accepted for publication).
9. Chitlur M, Sorensen B, Rivard GE, Young G, Ingerslev J, et al. (2011) Standardization of thromboelastography: a report from the TEG-ROTEM working group. *Haemophilia* 17: 532-537.
10. Kitchen DP, Kitchen S, Jennings I, Woods T, Walker I (2010) Quality assurance and quality control of thrombelastography and rotational Thromboelastometry: the UK NEQAS for blood coagulation experience. *Semin Thromb Hemost* 36: 757-763.

CHAPTER 5

Accuracy of the point-of-care coagulometer CoaguChek XS in the hands of patients

Michael Nagler, Pascale Raddatz Müller, Pirmin Schmid, Lucas M. Bachmann, Walter A. Willemin

Journal of Thrombosis and Haemostasis 2013, 11(1): 197-9

Dear Editor,

In a recent systematic review published in the Journal, Christensen et al. found no high-quality papers assessing the diagnostic value of point of care testing (POCT) coagulometers when used by patients, and they concluded that the question of accuracy remains open in this setting [1]. Stimulated by this excellent report, we present the results of an investigator-initiated, non-sponsored diagnostic study investigating the concordance of INR values determined by patients using the POCT coagulometer CoaguChek XS® (Roche Diagnostics, Basel, Switzerland) and determined by certified laboratories using an established reference method.

In this paper we defined precision as the extent of reproducibility of a test result (influenced by analytical precision, biological variation and pre-analytical factors) and accuracy as the extent of agreement between an index test and a reference test (analytical accuracy) or the true illness state (diagnostic accuracy) [1-4]. To assess whether diagnostic accuracy was sufficient, we applied the trade-offs proposed earlier [1,5].

We analysed data of all patients who had received training and were performing patient self-management (PSM) within the “Coagulationcare” initiative in Switzerland between 2006 and 2009 (about 90% of PSM patients in Switzerland) [6,7]. At the check-up visit one to three months after the PSM training, patients tested their INR value themselves. Second, venous blood samples were taken within one hour and the prothrombin time was determined in the laboratory of each training centre (the tertiary care hospitals of Lucerne, Berne, Basel and Zurich), using a one-stage clotting assay employing the reagent Innovin® (Siemens Healthcare Diagnostics AG, Zurich, Switzerland; ISI 0.96–0.99) on STA-R® coagulation analyser (Roche diagnostics AG, Rotkreuz, Switzerland), BCS® or BCS-XP® coagulation analyser (Siemens Healthcare Diagnostics AG, Zurich, Switzerland). Calibration to ISI standards was done with each batch of the Innovin® reagent using the InnoCal Set (Siemens Healthcare Diagnostics AG, Zurich, Switzerland). Standard measures of agreement were used to examine the extent of concordance [1,5,8]. Statistical analysis was performed with the MedCalc software (Version 12.2.1.0, Mariakerke, Belgium).

INR data on 534 out of 657 patients trained were available; 21 (3.2%) had not yet started PSM, and 102 (15.5%) did the check-up visit with the family physician. Median age was 54.1 years (range 16.2 – 85.1), and 169 were women (31.6%). Indications for oral anticoagulation (OAC; almost exclusively phenprocoumon) were: venous thromboembolism (n=219; 41.1%), prosthetic heart valve/aortic graft (n=154; 28.9%), atrial fibrillation (n=112; 21.0%), arterial thromboembolism (n=21; 3.9%) and others (n=28; 5.2%). All applied measures revealed a good level of agreement with regard to mean difference, limits of agreement, magnitude of difference, and concordance (**Figure 1**). Based on high-quality criteria established before [1,5] – a large number of measure-

ments, controlled laboratory as comparator, use of reputable accuracy criteria, not industry-sponsored – our investigation shows that the accuracy of CoaguChek XS is adequate for clinical use if used by patients to determine their INR value by themselves.

This report closes a relevant gap in evidence that has been identified in a recent review published in the Journal [1]. Our results are in line with previous investigations that found a good to acceptable, but somewhat inconsistent accuracy of CoaguChek XS [1,9-15] in the hands of medical staff. In conclusion, the CoaguChek XS is a reliable instrument in POCT and of value in patient self-management of oral anticoagulation.

| Accuracy criterion | Results |
|-------------------------|--------------------------------|
| Mean difference | -0.08 |
| Limits of agreement | -0.60 to +0.44 |
| Magnitude of difference | <i>Numbers of patients (%)</i> |
| 0 – 0.5 | 502 (94%) |
| 0.5 – 1.0 | 32 (6%) |
| > 1.0 | 0 |
| Concordance (narrow)* | 97.4% |
| Concordance (expanded)+ | 97.8% |

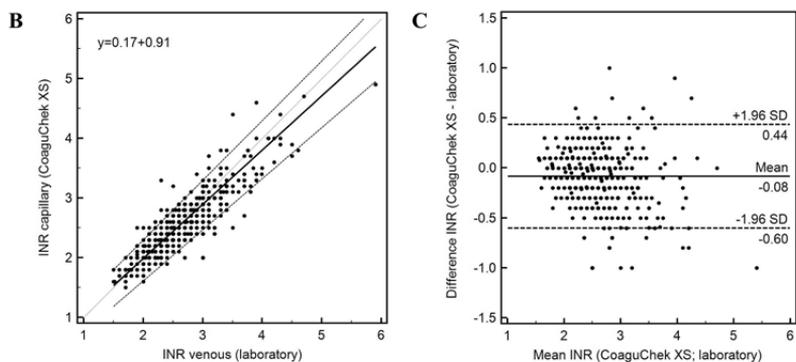


Figure 1. Accuracy of the POCT coagulometer CoaguChek[®] XS if used by patients to determine their INR values by themselves. (A) Accuracy criteria and corresponding results. Criteria for concordance (narrow) are fulfilled if both measurements are within the target range, or if both are above the range and the difference is maximal 0.8 units, or if both are below the range and the difference is maximal 0.4 units, or if one measurement is within the range and the difference is maximal 0.5 units (*). Criteria for concordance (expanded) are fulfilled if both measurements are within the target range, or if both are below the target range, or if both are above the target range, or if one measurement is within the target range and the difference is maximal 0,5 units (+). (B) Scatter diagram with Passing & Bablok regression line (solid line) and identity line (dotted line). (C) Bland-Altman difference plot. Horizontal lines are drawn at the mean differences and the limits of agreement (± 1.96 standard deviation).

References

1. Christensen TD, Larsen TB (2012) Precision and accuracy of point-of-care testing coagulometers used for self-testing and self-management of oral anticoagulation therapy. *J Thromb Haemost* 10: 251-260.
2. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 46: 1041-1050.
3. Gordis L (2008) Assessing the Validity and Reliability of Diagnostic and Screening Tests. Epidemiology. 4th ed. Philadelphia: Saunders Elsevier. pp. 85-108.
4. Banfi G, Del Fabbro M (2009) Biological variation in tests of hemostasis. *Semin Thromb Hemost* 35: 119-126.
5. Anderson DR, Harrison L, Hirsh J (1993) Evaluation of a portable prothrombin time monitor for home use by patients who require long-term oral anticoagulant therapy. *Arch Intern Med* 153: 1441-1447.
6. Nagler M, Schmid P, Raddatz-Muller P, Wuillemin WA (2012) Efficacy and safety of patient self-management of oral anticoagulation with vitamin K antagonists in everyday practice in a large population-based study in Switzerland. *Hämostaseologie* 32: A25.
7. Fritschi J, Raddatz-Muller P, Schmid P, Wuillemin WA (2007) Patient self-management of long-term oral anticoagulation in Switzerland. *Swiss Med Wkly* 137: 252-258.
8. Douketis JD, Lane A, Milne J, Ginsberg JS (1998) Accuracy of a portable International Normalization Ratio monitor in outpatients receiving long-term oral anticoagulant therapy: comparison with a laboratory reference standard using clinically relevant criteria for agreement. *Thromb Res* 92: 11-17.
9. Bauman ME, Black KL, Massicotte MP, Bauman ML, Kuhle S, et al. (2008) Accuracy of the CoaguChek XS for point-of-care international normalized ratio (INR) measurement in children requiring warfarin. *Thromb Haemost* 99: 1097-1103.
10. Braun S, Watzke H, Hasenkam JM, Schwab M, Wolf T, et al. (2007) Performance evaluation of the new CoaguChek XS system compared with the established CoaguChek system by patients experienced in INR-self management. *Thromb Haemost* 97: 310-314.
11. Christensen TD, Larsen TB, Jensen C, Maegaard M, Sorensen B (2009) International normalised ratio (INR) measured on the CoaguChek S and XS compared with the laboratory for determination of precision and accuracy. *Thromb Haemost* 101: 563-569.
12. Greenway A, Ignjatovic V, Summerhayes R, Newall F, Burgess J, et al. (2009) Point-of-care monitoring of oral anticoagulation therapy in children. Comparison of the CoaguChek XS system with venous INR and venous INR using an International Reference Thromboplastin preparation (rTF/95). *Thromb Haemost* 102: 159-165.
13. Paioni P, Kroiss S, Kagi E, Bergstrasser E, Fasnacht M, et al. (2009) Self-monitoring of oral anticoagulation therapy in children. *Acta haematologica* 122: 58-63.
14. Sobieraj-Teague M, Daniel D, Farrelly B, Coghlan D, Gallus A (2009) Accuracy and clinical usefulness of the CoaguChek S and XS Point of Care devices when starting warfarin in a hospital outreach setting. *Thromb Res* 123: 909-913.
15. Williams VK, Griffiths AB (2007) Acceptability of CoaguChek S and CoaguChek XS generated international normalised ratios against a laboratory standard in a paediatric setting. *Pathology* 39: 575-579.

CHAPTER 6

Prospective evaluation of the interobserver reliability of the 4T's score in patients with suspected heparin-induced thrombocytopenia

Michael Nagler, Thomas Fabbro, Walter A. Wuillemin

Published in part: *Journal of Thrombosis and Haemostasis* 2012; 10: 151–2.

Abstract

Background: Suspicion of heparin-induced thrombocytopenia (HIT) necessitates an immediate diagnostic workup to prevent severe complications. Diagnostic strategy includes clinical assessment and 4Ts score is the most common score. The interobserver reproducibility of the 4Ts score in real-life clinical practice is unknown.

Objectives: To prospectively investigate the interobserver reliability of the 4Ts score in a real-life setting.

Methods: In 40 consecutive patients with suspected HIT the 4Ts score was rated independently by the attending physician, a haematologist and the principle investigator. Raw agreement was used to estimate the agreement regarding the three risk categories and overall as well as for individual items of the 4Ts score. A generalised estimating equation was used to analyse a systematic shift between professions.

Results: Agreement observed for category low risk was 48% (95% CI: 25-68%), for intermediate risk 65% (95% CI: 52-79%) and for high risk 47% (95% CI: 22-68%). Overall an agreement of 58% (95% CI: 45-69%) was found. Regarding the individual items agreement was 78% (95% CI: 68-89%) for T1 (magnitude of thrombocytopenia), 55% (95% CI: 43-66%) for T2 (timing of thrombocytopenia), 89% (95% CI: 81-96%) for T3 (presence of thrombosis) and 62% (95% CI: 50-73%) for T4 (existence of other causes for thrombocytopenia). The average ratings of the attending physicians, haematologists and the principle investigator were 4.46, 4.65 and 4.15 respectively and differed significantly ($p=0.026$).

Conclusions: Our investigation indicates that interobserver reliability of the 4Ts score is limited in a real-life setting. Determination of the items of the 4Ts score should be further specified.

Background

Suspected heparin-induced thrombocytopenia (HIT) needs an immediate diagnostic workup to prevent severe complications [1]. HIT is regarded as a clinicopathologic syndrome, and clinical characteristics as well as laboratory findings have to be considered for the diagnosis [1,2]. Platelet factor 4 (PF4) complexed to heparin form the antigen for the development of IgG antibodies, which can be recognized in up to 50% of the patients treated with heparin [3-5]. In some patients circulating immune complexes boost platelet activation and generation of thrombin leading to extensive venous and arterial thrombosis, thrombocytopenia, skin reactions, limb loss and death [6]. Not only a delay in diagnosis, but overtreatment sets patients at risk because therapy with alternative anticoagulants is associated with a high rate of bleeding complications, severe anaphylactic complications and a worse manageability than heparins [7-10].

The diagnostic strategy must comprise a clinical pre-test scoring system to increase the low specificity of high sensitive antibody tests and the 4T's score is probably the score most often used in this context [1]. Recent studies questioned the inter-observer reliability of the 4T's score [11,12]. However, determination of 4T's scores in previous studies does not reflect real-life clinical practice. In these studies the score had been usually calculated retrospectively, by laboratory staff receiving specimens for antibody testing, by pharmacists, by specially trained physicians or by investigators familiar with HIT. Since no prospective evaluation in a real-life setting exists, the applicability of results of validation studies in clinical practice may be limited. To mention this issue, we prospectively investigated the inter-observer reliability of the 4T's score in patients with suspected HIT in a real-life setting.

Methods

Patients and study design

This study was a prospective observational study. From June 2010 to March 2011, 40 consecutive patients evaluated for suspected HIT were included if they met the following criteria: expected in-patient period of more than 3 days and no evaluation for suspected HIT during the last month. Patient characteristics and clinical information regarding 4T's score and evaluation of suspected HIT were collected prospectively by chart review and discussion with the ward physician. The study protocol has been approved by the local ethical review board of our institution (Kantonale Ethikkommission Luzern).

Assessment of the 4Ts score

Three physicians rated the 4T's score independently for every patient: (i) the attending physician who asked for antibody testing (n = 36; a specialist registrar in intensive care medicine, internal medicine or general surgery), (ii) the haematologist on duty (n = 6) and (iii) the principal investigator (PI; M.N.). The clinical and laboratory data were obtained by chart review and by visiting the patient, whilst the score's calculation was blinded between observers. Determination of the score was done before antibody testing by attending physician and haematologist and usually on the same day by the PI. Only data before the time of antibody testing were considered, results of antibody testing were not known. A structured, standardised questionnaire was developed taking previous published details and specifications into account (see supplementary material S1) [13,14]. This questionnaire was incorporated into the hospital's local patient management. Before starting the study a training lecture was given to haematology staff members. Attending physicians were supported in case of questions regarding particular items of the 4T's score.

Laboratory diagnostics

HIT antibody testing was performed using the particle-gel immune assay ID-H/PF4-PaGIA (DiaMed SA, Cressier sur Morat, Switzerland) [15,16]. Patient plasma was added to the reaction chamber, incubated and centrifuged according to the manufacturers' instructions. PaGIA was defined as positive if red polymers remained on top of the reaction chamber. In case of undetermined results, serial dilution was performed [17]. Platelet count was measured with XE 5000 (Sysmex AG, Horgen, Switzerland).

Data analysis

To quantify the inter-rater reliability between different observers of the 4T's score we used the raw agreement because it can easily be interpreted. The raw agreement describes the proportion of possible agreement realised. We calculated the raw agreement for the three individual risk categories low risk (4T's score: 0-3 points), intermediate risk (4T's score: 4-5 points), and high risk (4T's score: 6-8 points) as well as for the overall risk category rating. Additionally we calculated the raw agreement for the individual items of the 4T's score (T1 = magnitude of thrombocytopenia, T2 = timing of thrombocytopenia, T3 = presence of thrombosis, T4 = existence of other causes for thrombocytopenia) and for the total 4T's score. The 95% confidence intervals, CI, for raw agreement were calculated by parametric bootstrapping. For comparison with other publications, weighted Light's kappa was calculated for the 4T's score and its items (T1-T4). In the second part of the study, a systematic shift between the three observers, attending physician, haematologist and principle investigator, was analysed. Therefore, a generalised estimating equation was fit, that takes the relationship of the

three ratings per patient into account. All statistical analyses were performed and all plots constructed using the statistical software R (version 2.11) [18].

Results

Patient characteristics, clinical probability and laboratory results

Patient characteristics and results of antibody testing of the 40 patients included in the study are summarized in **Table 1**. In 20 percent of the patients a positive antibody testing (ID-H/PF4-PaGIA) was noticed. Clinical probability for presence of heparin-induced thrombocytopenia was different if rated by attending physician, haematologist or principle investigator. Distribution of the patients regarding clinical probability rated by the different physicians is given in **Table 2**.

Table 1: Characteristics of patients with suspected HIT (n=40).

| | |
|-------------------------|---------------------------|
| Age (years) | 63.4 (14-86) [*] |
| Female gender (n) | 23 |
| Medical patient (n) | 24 |
| Surgical patient (n) | 16 |
| Intensive care unit (n) | 21 |
| UFH (n) | 22 [†] |
| LMWH (n) | 29 [†] |
| positive PF4-PaGIA (n) | 8 |

^{*} Median (range). [†] 20% of the patients received both, UFH and LMWH.

Table 2: Distribution of the patients regarding clinical probability rated by attending physician, haematologist and principle investigator (n=40).

| | Attending physician | Haematologist | Principle Investigator |
|-----------------------|---------------------|---------------|------------------------|
| Low risk (%) | 52.5 | 40 | 70 |
| Intermediate risk (%) | 42.5 | 52.5 | 22.5 |
| High risk (%) | 5 | 7.5 | 7.5 |

Agreement of the risk classification

The raw agreement of the different observers is shown in **Figure 1**. We observed a very limited agreement for all three categories of the 4T's score. A 48% agreement (95% CI: 25-68%) was observed for category low risk, a 65% (95% CI: 52-79%) agreement for

category intermediate risk and a 47% (95% CI: 22-68%) agreement for category high risk. Overall agreement was 58% (95% CI: 45-69%).

The agreement of the individual items of the 4T's score is indicated in **Figure 2**. Agreement for item T1 was 78% (95% CI: 68-89%), for item T2 55% (95% CI: 43-66%), for item T3 89% (95% CI: 81-96%) and for item T4 62% (95% CI: 50-73%). Overall a raw agreement of 34% (95% CI: 22-43%) was observed.

For comparison with other publications, weighted Light's kappa was calculated as for T1 0.733 (95% CI: 0.536-0.877), T2 0.451 (95% CI: 0.256-0.620), T3 0.705 (95% CI: 0.425-0.919), T4 0.234 (95% CI: 0.078-0.426), and for the overall score 0.614 (95% CI: 0.421-0.815).

Systematic shift between professions

The ratings of the different observers are summarised in **Figure 3**. The average ratings were estimated as 4.46, 4.65, 4.15, for attending physicians, haematologists and the principle investigator, respectively, and differed significantly (p=0.026).

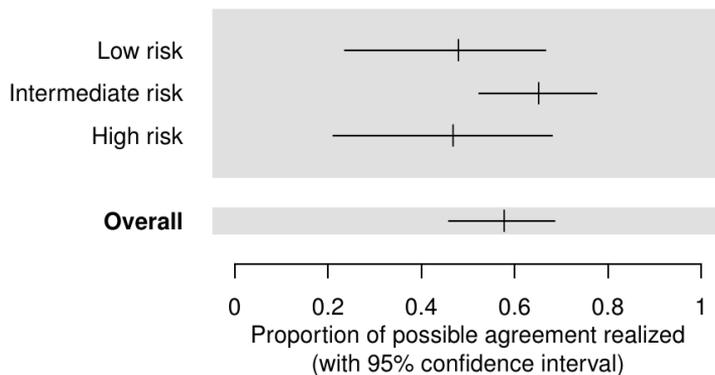


Figure 1: Raw agreement of the classification of patients into low, intermediate and high risk.

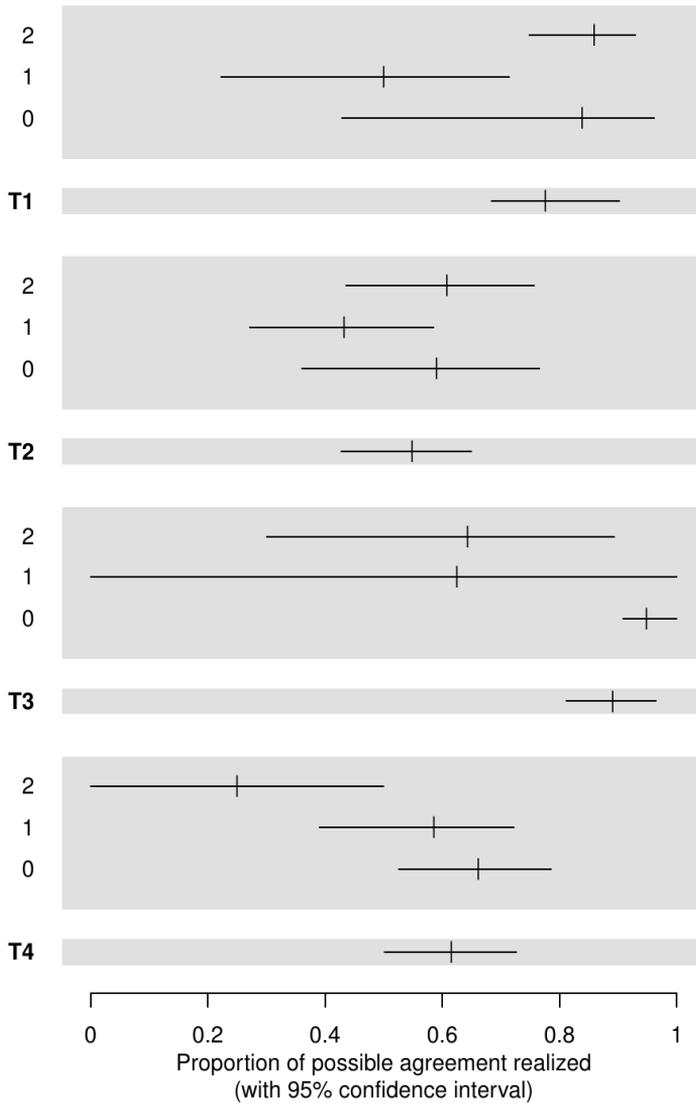


Figure 2: Raw agreement for the four items of the 4T's score (T1-T4) and their levels.

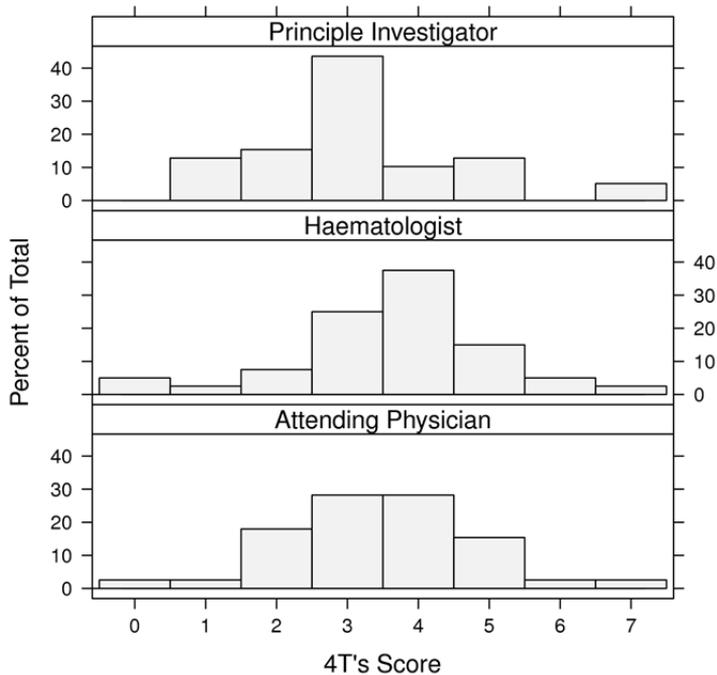


Figure 3: Histograms of the 4T's score rated by principle investigator, haematologist and attending physician

Discussion

We observed a very limited agreement of the different observers for all three categories of the 4Ts score. Our data indicate, that the interobserver reliability of the 4Ts score is limited in a real-life setting. Discrepancies arise particularly for T2 (timing of thrombocytopenia) and T4 (existence of other causes of thrombocytopenia). Furthermore, there were small but significant differences when the 4Ts score was rated by an attending physician, a hematologist, or the PI.

Although previous investigations have not primarily addressed reliability in real-life situations, our observations are in line with them with the lower agreement between different observers. In a recent investigation an intraclass correlation coefficient of $r = 0.71$ (95% CI 0.54 – 0.83) was found [11]. In another investigation discrepancies in up to 50% of the patients were described [12]. Even Lo and colleagues proposed problems of inter-observer agreement as a possible reason for variable results in evaluation of 4T's score [13].

The main limitation of our study is, that no functional antibody testing has been performed. Thus, we don't know if there's an observer group that rated more correct than the other. However, the aim of our investigation was to show the differences and

not which observer was more accurate. But, some of the observers are more experienced than the others and our investigation revealed that levels of insight influence scoring of the 4T's score. The strength of our investigation is that it was done in real-life clinical practice. In previous publications determination of the 4T's score was whether done retrospectively, exclusively by chart review or by specialised staff different from treating physicians [11-13]. Moreover, the authors of these publications themselves stated this aspect as a limitation of their investigations.

A remarkable low agreement was observed regarding T2 (timing of thrombocytopenia; 55%; 95% CI 43-66%) and T4 (existence of other causes of thrombocytopenia; 62%; 95% CI 50-73%). This reflects our experience of daily clinical practice. Most discussions are about time of onset of thrombocytopenia and possible other causes. Discrepancies in these items were also noticed in an observational trial discussed above [12]. These results indicate, that there's an urgent need for specifications of the 4Ts score.

A systematic difference of the scoring was seen between the observer groups. The haematologists on duty scored systematically higher and the PI lower than the attending physicians. This phenomenon is perhaps reflection of the duty haematologist(s) being alert not to miss the diagnosis of HIT, whilst the PI might systematically avoid over-diagnosis of HIT, amongst other things to prevent unnecessary side effects of alternative anticoagulants. Due to the design of the study, there are no data, which enable correctness of observer judgements to be compared.

In conclusion, our study shows that inter-observer reliability of 4T's score is limited in a real-life setting. Discrepancies arise in particularly in case of T2 (timing of thrombocytopenia) and T4 (existence of other causes for thrombocytopenia). Furthermore, there are small but significant differences if the 4T's score was rated by an attending physician, a haematologist or the PI, suggesting that even with a standardised approach real-life differences in awareness of the complex aspects of HIT undermine the 4T's score's reliability. This may limit application of results of validation studies of 4T's score to clinical practice. Determination of the items of the 4T's score should be further specified as suggested by Warkentin and colleagues [14] to overcome this limitation. Furthermore, evaluation of inter-observer agreement in real-life settings is a prerequisite for developing future clinical assessment tools.

References

1. Greinacher A (2009) Heparin-induced thrombocytopenia. *Journal of thrombosis and haemostasis: JTH 7 Suppl 1*: 9-12.
2. Warkentin TE, Greinacher A, Koster A, Lincoff AM (2008) Treatment and prevention of heparin-induced thrombocytopenia: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 340S-380S.
3. Rauova L, Arepally G, McKenzie SE, Konkle BA, Cines DB, et al. (2009) Platelet and monocyte antigenic complexes in the pathogenesis of heparin-induced thrombocytopenia (HIT). *J Thromb Haemost 7 Suppl 1*: 249-252.
4. Warkentin TE, Cook RJ, Marder VJ, Greinacher A (2010) Anti-PF4/heparin antibody formation postorthopedic surgery thromboprophylaxis: the role of non-drug risk factors and evidence for a stoichiometry-based model of immunization. *J Thromb Haemost 8*: 504-512.
5. Warkentin TE, Sheppard JA, Horsewood P, Simpson PJ, Moore JC, et al. (2000) Impact of the patient population on the risk for heparin-induced thrombocytopenia. *Blood* 96: 1703-1708.
6. Warkentin TE, Kelton JG (1996) A 14-year study of heparin-induced thrombocytopenia. *Am J Med* 101: 502-507.
7. Greinacher A, Lubenow N, Eichler P (2003) Anaphylactic and anaphylactoid reactions associated with lepirudin in patients with heparin-induced thrombocytopenia. *Circulation* 108: 2062-2065.
8. Greinacher A, Eichler P, Lubenow N, Kwasny H, Luz M (2000) Heparin-induced thrombocytopenia with thromboembolic complications: meta-analysis of 2 prospective trials to assess the value of parenteral treatment with lepirudin and its therapeutic aPTT range. *Blood* 96: 846-851.
9. Tardy B, Lecompte T, Boelhen F, Tardy-Poncet B, Elalamy I, et al. (2006) Predictive factors for thrombosis and major bleeding in an observational study in 181 patients with heparin-induced thrombocytopenia treated with lepirudin. *Blood* 108: 1492-1496.
10. Lubenow N, Eichler P, Lietz T, Farner B, Greinacher A (2004) Lepirudin for prophylaxis of thrombosis in patients with acute isolated heparin-induced thrombocytopenia: an analysis of 3 prospective studies. *Blood* 104: 3072-3077.
11. Cuker A, Arepally G, Crowther MA, Rice L, Datko F, et al. (2010) The HIT Expert Probability (HEP) Score: a novel pre-test probability model for heparin-induced thrombocytopenia based on broad expert opinion. *J Thromb Haemost 8*: 2642-2650.
12. Strutt JK, Mackey JE, Johnson SM, Sylvia LM (2011) Assessment of the 4Ts pretest clinical scoring system as a predictor of heparin-induced thrombocytopenia. *Pharmacotherapy* 31: 138-145.
13. Lo GK, Juhl D, Warkentin TE, Sigouin CS, Eichler P, et al. (2006) Evaluation of pretest clinical score (4 T's) for the diagnosis of heparin-induced thrombocytopenia in two clinical settings. *J Thromb Haemost 4*: 759-765.
14. Warkentin TE, Linkins LA (2010) Non-necrotizing heparin-induced skin lesions and the 4T's score. *J Thromb Haemost 8*: 1483-1485.
15. Pouplard C, Gueret P, Fouassier M, Ternisien C, Trossaert M, et al. (2007) Prospective evaluation of the '4Ts' score and particle gel immunoassay specific to heparin/PF4 for the diagnosis of heparin-induced thrombocytopenia. *J Thromb Haemost 5*: 1373-1379.
16. Meyer O, Salama A, Pittet N, Schwind P (1999) Rapid detection of heparin-induced platelet antibodies with particle gel immunoassay (ID-HPF4). *Lancet* 354: 1525-1526.
17. Alberio L, Kimmerle S, Baumann A, Taleghani BM, Biasiutti FD, et al. (2003) Rapid determination of anti-heparin/platelet factor 4 antibody titers in the diagnosis of heparin-induced thrombocytopenia. *The American journal of Medicine* 114: 528-536.
18. R Development Core Team (2011) *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

APPENDIX

Diagnostic value of the 4T's score for heparin-induced thrombocytopenia in the critically ill

Comment on: M. Crowther et. al. J Crit Care. 2014 Jun; 29(3):470.e7-470.e15

Michael Nagler, Anne Angelillo-Scherrer

Journal of Critical Care. 2014; 29(6): 1126-27

Dear Editor,

With great interest, we read the article by Crowther et al [1] and the corresponding editorial by McMahon et al [2], both recently published in the *Journal of Critical Care* and focusing on the diagnostic value of a clinical scoring system (the 4Ts score) [3] for heparin-induced thrombocytopenia (HIT) in critically ill patients. We congratulate Crowther and colleagues for their excellent work but we would like to raise awareness for two issues that remained incompletely addressed in their report.

First, we would like to emphasize that the results of this study, with a low prevalence of HIT of less than one percent should not be thoughtlessly transferred to settings, where the 4Ts score is typically applied. Previous data suggest that the prevalence of HIT in these settings is as high as fifteen percent [3,4]. The low prevalence of HIT makes the 4Ts score to appear better than what can be realistically expected in clinical practice. Obviously, ruling-out a condition that is very uncommon in a study sample is not challenging. Unsurprisingly, the 4Ts score had a high negative predictive value in the paper at hand (our **Table 1**).

To illustrate our point, we calculated the positive and negative likelihood ratios using the data from the two-by-two table (our **Table 2**). Applying Bayes theorem, the **Figure 1** shows the pretest probability and corresponding posttest probabilities given a positive or negative test result. For pretest probabilities between ten and twenty percent, the related posttest probabilities would range from eight to twenty-two percent, when the test is negative. Conversely, posttest probabilities given a positive test result would yield values between thirty and sixty percent. For positive and negative test results the remaining uncertainty appears to be unacceptably high to rule-in or rule-out HIT.

Second, some methodological issues deserve being mentioned. Patients were drawn from a randomized controlled trial that excluded several patients at risk for HIT (“heparin administration of at least 3 days”; “need for therapeutic anticoagulation”).

Moreover, the nested cohort excluded some patients at risk without applying pre-specified reasons, missed at least four patients with a positive reference standard test and excluded patients with venous thromboembolism. Furthermore, the inclusion criteria of the nested study cohort comprised already several items of the 4Ts score (decrease of platelet count (PLT) to less than $50 \times 10^9/L$; PLT decrease more than 50%; development of VTE) and important information required to determine the 4Ts score was censored (heparin treatment before the admission to the ICU). Moreover, the reference standard was not accurately specified and some patients did not undergo reference standard testing. For example, the serotonin release assay (SRA) was inconsistently applied and complete data on diagnostic tests were available for 432 out of 794 patients only. Finally, experts or specially trained investigators determined the 4Ts score which leads to a higher reproducibility of 4Ts score results than what can be expected in real-life clinical practice [5], [3,6].

In conclusion, we agree with the authors that well-planned, prospective investigations are still necessary. The data suggest that further studies should focus on the development and validation of new tests and diagnostic algorithms, combining not only clinical information but also results of rapidly available laboratory assays.

Table 1: Accuracy criteria

| Criterion | Measure |
|---------------------------|-------------------------------|
| Sensitivity | 45.5% (95% CI 16.8% to 76.7%) |
| Specificity | 86.6% (95% CI 83.2% to 89.6%) |
| Positive predictive value | 7.5% (95% CI 2.5% to 16.6%) |
| Negative predictive value | 98.5% (95% CI 96.8 to 99.5%) |

Table 2: Two-by-two contingency table

| 4Ts score | Diagnosis | | Total |
|------------------------|-----------|--------|-------|
| | HIT | No HIT | |
| Intermediate/high risk | 5 | 62 | 67 |
| Low risk | 6 | 401 | 407 |
| Total | 11 | 463 | |

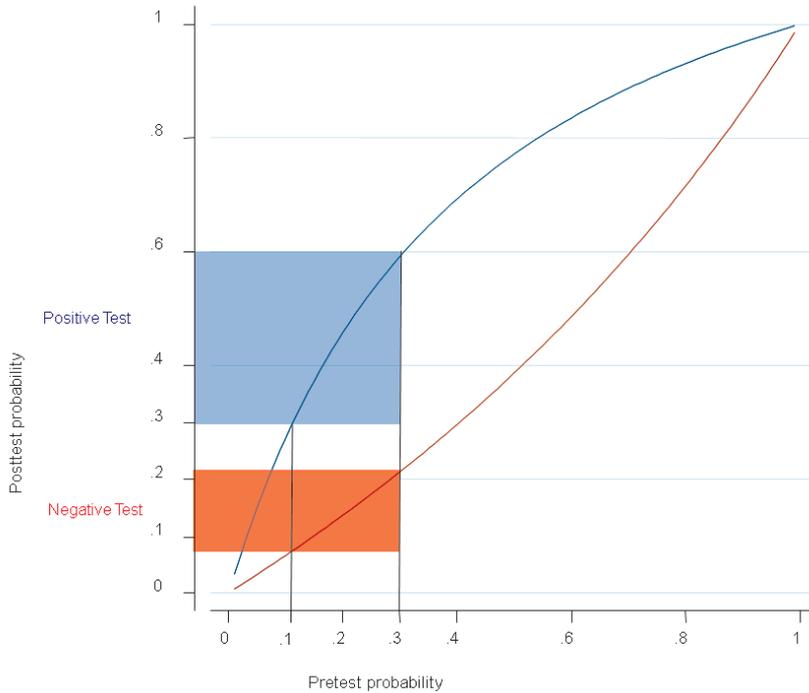


Figure 1: Diagnostic value of the 4Ts score for heparin-induced thrombocytopenia in ICU patients according to the prevalence. Post-test probabilities are shown for negative (low risk) and positive test results (intermediate and high risk).

References

1. Crowther M, Cook D, Guyatt G, Zytaruk N, McDonald E, et al. (2014) Heparin-induced thrombocytopenia in the critically ill: Interpreting the 4Ts test in a randomized trial. *J Crit Care* 29: 470 e477-470 e415.
2. McMahon CM, Cuker A (2014) The 4Ts test in the critically ill: Timing is everything. *J Crit Care* 29: 468-469.
3. Cuker A, Gimotty PA, Crowther MA, Warkentin TE (2012) Predictive value of the 4Ts scoring system for heparin-induced thrombocytopenia: a systematic review and meta-analysis. *Blood* 120: 4160-4167.
4. Nellen V, Sulzer I, Barizzi G, Lammler B, Alberio L (2012) Rapid exclusion or confirmation of heparin-induced thrombocytopenia: a single-center experience with 1,291 patients. *Haematologica* 97: 89-97.
5. Nagler M, Fabbro T, Wuillemin WA (2012) Prospective Evaluation of the Inter-Observer Reliability of the 4T's Score in Patients with Suspected HIT. *Journal of thrombosis and haemostasis: JTH* 10: 151-152.
6. Lo GK, Juhl D, Warkentin TE, Sigouin CS, Eichler P, et al. (2006) Evaluation of pretest clinical score (4 T's) for the diagnosis of heparin-induced thrombocytopenia in two clinical settings. *J Thromb Haemost* 4: 759-765.

CHAPTER 7

Variability between laboratories performing coagulation tests with identical platforms: a nationwide evaluation study

Michael Nagler, Lucas M. Bachmann, Lorenzo Alberio, Anne Angelillo-Scherrer, Lars M. Asmis, Wolfgang Korte, Adriana Mendez, Guido Reber, Hans Stricker, Dimitrios A. Tsakiris, Walter A. Wuillemin

Thrombosis Journal 2013, 11:6

Abstract

Background: While the assessment of analytical precision within medical laboratories has received much attention in scientific enquiry, the degree of as well as the sources causing variation between them remains incompletely understood. In this study, we quantified the variance components when performing coagulation tests with identical analytical platforms in different laboratories and computed intraclass correlations coefficients (ICC) for each coagulation test.

Methods: Data from eight laboratories measuring fibrinogen twice in twenty healthy subjects with one out of 3 different platforms and single measurements of PT, and coagulation factors II, V, VII, VIII, IX, X, XI and XIII were analysed. By platform, the variance components of (i) the subjects, (ii) the laboratory and the technician and (iii) the total variance were obtained for fibrinogen as well as (i) and (iii) for the remaining factors using ANOVA.

Results: The variability for fibrinogen measurements within a laboratory ranged from 0.02 to 0.04, the variability between laboratories ranged from 0.006 to 0.097. The ICC for fibrinogen ranged from 0.37 to 0.66 and from 0.19 to 0.80 for PT between the platforms. For the remaining factors the ICC's ranged from 0.04 (F II) to 0.93 (F VIII).

Conclusions: Variance components that could be attributed to technicians or laboratory procedures were substantial, led to disappointingly low intraclass correlation coefficients for several factors and were pronounced for some of the platforms. Our findings call for sustained efforts to raise the level of standardization of structures and procedures involved in the quantification of coagulation factors.

Introduction

While the assessment of analytical precision within one medical laboratory has received much attention in scientific enquiry, extent and sources of variability that occur between laboratories quantifying the same parameter remains incompletely understood. In view of the fact that the clinical value of a laboratory test depends directly on its reproducibility and comparability [1], the scientific community has made great efforts to promote analytical precision within laboratories lately [2-4]. For example, external quality assessment programs (EQA) were introduced to improve comparability between laboratories and are seen as an essential part of quality management systems today [3-5]. Nevertheless, data on variability between laboratories remain limited and analysis of variance components are scarce. Available investigations indicate a large variability, even in the context of coagulation parameter measurements [5-8] thus jeopardizing the comparability of results between different institutions.

On the other hand, knowledge is also scarce for causal sources of variability between laboratories. Possible influences may be the type of the parameter, reagents and calibrators used, the level of standardisation within each laboratory and the adherence to guidelines [9]. If investigations could identify factors that contribute to the variability between laboratories, efforts could be made to improve comparability efficiently. In this study, using the example of haemostasis, we quantified the variance components when performing coagulation tests with identical analytical platforms in different laboratories and computed intraclass correlations coefficients (ICC) for each coagulation test. Thus, we aimed to quantify the extend of variation between state-of-the-art laboratories in relation to the variation of the subjects and to determine the sources of this variation.

Materials and Methods

Study design and population

This investigation was a secondary analysis based on a previous study of the Swiss RIVAMOS study group [10]. In this multicentre, prospective evaluation study blood samples of 20 eligible and consenting healthy male volunteers were included. The study was approved by the local ethical review board of our institution (Kantonale Ethikkommission Luzern).

Variance components of coagulation tests between laboratories

Variability of coagulation tests between laboratories is influenced by biological and analytical variability. The two main components of biological variability are intra- and inter-individual variability [11]. Analytical variability between laboratories is influenced (a) by the variability within laboratories, (b) by the platforms used, (c) the assay designs

used and (d) variance components that could be attributable to technicians and laboratory procedures (including the choice of reagents and calibrators). The type of coagulation factor measured may also influence the variability. To investigate the variability between laboratories, we included the inter-individual variability, the variability within laboratories, the variability between laboratories and the remaining variability (variance components, that could be attributable to technicians and laboratory procedures) in our statistic model as described below. The remaining factors are considered as stable due to several reasons. The intra-individual variability is regarded as low in coagulation tests and the interval between the two measurements is very short (2-3 hours) [11]. Influence of the platform was addressed as calculations have been done by every platform. The assay designs were identical by platform. Preanalytic factors were considered by applying a strict protocol and using the same technician for serial measurements. Storage and transport of the samples were identical for all laboratories. To avoid influences of different batches of reagents, all determinations were done in one batch in every laboratory.

Blood withdrawal and handling of the samples

Citrated plasma samples were taken two times in an interval of 2 to 3 hours at the senior author's institution. Blood withdrawal was done under standardised conditions to preclude preanalytic influences. Plastic syringes (Monovette[®], Sarstedt, Nümbrecht, Germany) containing 1 ml trisodium citrate (0.106 mol/l) for 9 ml of blood were used. Citrated samples were collected after EDTA samples to avoid tissue factor contamination of citrated samples. Plasma samples were snap-frozen at – 20°C and shipped under standardised conditions using a commercial courier service with a delivery time of 2-4 hours.

Laboratories and analysis

All nine haemostasiological laboratories of tertiary hospitals in Switzerland participated in this investigation (In alphabetical order; Cantonal Hospital Aarau, University Hospital of Basel, Inselspital University Hospital of Berne, University Hospital of Geneva, University Hospital of Lausanne, Regional Hospital of Locarno, Cantonal Hospital of Lucerne, Cantonal Hospital of St. Gallen, University Hospital of Zurich). One laboratory was excluded from analysis of between-laboratory variability because of a unique platform. The remaining eight laboratories used one out of three platforms. Details of coagulometers, assays and reagents used are given in **Table 1**. For determination of variability between laboratories, local protocols were used for testing samples. All laboratories used strict internal and external quality assessment measures, conducted formal test evaluation and are accredited to the Swiss Accreditation Service (SAS). Coagulation tests have been done in accordance with international guidelines [12].

Table 1. Participating laboratories with coagulometers, assays and reagents used

| Laboratory [¶] | Platform | Extrinsic coagulation factors [‡] | | Intrinsic coagulation factors [§] | | Fibrinogen | | Factor XIII | |
|-------------------------|--|--|------------------------------|--|------------------------------|---------------|-----------------------------------|---------------------------|------------------------------|
| | | Assay | Reagents | Assay | Reagents | Assay | Reagents | Assay | Reagents |
| A [†] | Instrumentation Laboratory ACL Top 700 | Clotting test (photometric) | Recombi-plastin [*] | Clotting test (photometric) | Hemosil [®] APPT SP | Clauss method | Hemosil [®] Fibrinogen-C | Kinetic photometric assay | Berichrom [®] FXIII |
| B | Siemens BCS-XP | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Actin [*] | Clauss method | Multifibren U | Kinetic photometric assay | Berichrom [®] FXIII |
| C | Roche STA-R | Clotting test (viscosity) | Innovin [*] | Clotting test (viscosity) | Actin FS [*] | Clauss method | STA [®] Fibrinogen | Latex immunturbidimetry | HEXAMATE [®] FXIII |
| D | Symex CA7000 | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Actin [*] | Clauss method | Dade [®] Thrombin | n.a. | n.a. |
| E | Siemens BCS-XP | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Pathromtin [*] SL | Clauss method | Multifibren U | Kinetic photometric assay | Berichrom [®] FXIII |
| F | Siemens BCS | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Actin FS [*] | Clauss method | Multifibren U | Kinetic photometric assay | Berichrom [®] FXIII |
| G [*] | Symex CA1500 | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Actin FS [*] | Clauss method | Dade [®] Thrombin | n.a. | n.a. |
| H | Siemens BCS-XP | Clotting test (photometric) | Innovin [*] | Clotting test (photometric) | Pathromtin [*] SL | Clauss method | Multifibren U | Kinetic photometric assay | Berichrom [®] FXIII |
| I | Roche STA-R | Clotting test (viscosity) | Innovin [*] | Clotting test (viscosity) | STA [*] APPT | Clauss method | bovine thrombin, Diagnostec AG | Kinetic photometric assay | Berichrom [®] FXIII |

[¶] Random order, not matching with the alphabetical list mentioned above. ^{*} Laboratory was excluded from analysis of variability between laboratories because not all tests were available. [†] Laboratory was excluded from analysis of variability between laboratories because of a unique platform. [‡] PT, Factors II, V, VII, X. [§] Factors VIII, IX, XI.

Fibrinogen was determined in plasma samples of both points in time. Prothrombin time (PT), factors II, V, VII, VIII, IX, X, XI and XIII were determined only for the first sample to avoid possible interferences with the second investigation conducted [10].

Statistical analysis

To study the variability between the different laboratories, we performed analyses of variance (ANOVAs) and used the variance components procedure. By platform, the variance components of (i) the subjects, (ii) the laboratory and the technician (including the measurement error of the platform) and (iii) the total variance were obtained for fibrinogen as well as (i) and (iii) for the remaining factors.

Based on the variance components, we calculated the intraclass correlation coefficients (ICC) for all factors using the variance component of the subjects in the numerator and the total variance in the denominator. Thus, the variation of the measurements between laboratories can be demonstrated in relation to the variation of the subjects (what represents the 'real' variation). A high ICC (clearly above 0.5) refers to the subjects as main components of the variance (which is a desirable outcome) whereas a low ICC indicates that other (non-subject) factors are more accountable (which is a disturbing outcome).

Statistical analysis was performed using the SPSS statistical software package (Version 19; Property of SPSS Inc an IBM Company. © Copyright SPSS Inc. 1989, 2010).

Results

A total of 360 fibrinogen measurements, 180 PT measurements, 160 F II, F V, F VII, F VIII, F IX, F X and F XI measurements, and 140 F XIII measurements were available for analysis because only fibrinogen was determined two times and one laboratory did not provided all tests.

The intraobserver variability for fibrinogen measurements, indicating the variability within a laboratory, ranged from 0.02 to 0.04. This variability occurs from the variability within a laboratory (including the technician handling the samples). The variability between laboratories ranged from 0.006 to 0.097. This indicates that laboratories used the platform in different ways or that they interpreted the manual differently. This detailed analysis was only possible for the fibrinogen measurement because fibrinogen was determined in plasma samples of both points in time.

The results allow discussing how prone a platform – the specifications, the complexity in the handling and other reasons – might be to create higher variability when used in different laboratories and contexts.

The ICC for fibrinogen ranged from 0.37 to 0.66 and from 0.19 to 0.80 for PT between the laboratories. Though, the amount to which non-subject factors is accountable for

the variance of the parameters is very varying. Low ICC indicates that non-subject factors are major components of variance.

The ICC of FII when assessed with platform 3 was as low as 0.04 as compared to 0.44 when measured with platform 2. For the remaining factors the ICC's ranged from 0.22 (F V) assessed with platform 3, to 0.93 (F VIII) assessed with platform 2.

For the two platforms providing enough data, the average ICC across all parameters was 0.64 (standard deviation 0.25) for platform 2 and 0.56 (standard deviation 0.28) for platform 3. Overall, the ICC's were at 0.60 on average. Details are available on **Table 2**.

Table 2: Intraclass Correlations of various coagulation tests and Intraobserver variability of fibrinogen tests with various platforms

| Parameter | Platform 1* | Platform 2 | Platform 3 | Platform 4 [‡] |
|----------------------------------|-------------|------------|------------|-------------------------|
| # laboratories | 1 | 2 | 4 | 2 |
| Fibrinogen | 0.85 | 0.37 | 0.65 | 0.66 |
| Within-laboratory variability ¶ | | 0.02 | 0.03 | 0.04 |
| Between-laboratory variability + | | 0.097 | 0.03 | 0.006 |
| Prothrombin time | | 0.19 | 0.24 | 0.80 |
| F II | | 0.44 | 0.04 | |
| F V | | 0.63 | 0.22 | |
| F VII | | 0.71 | 0.82 | |
| F VIII | | 0.93 | 0.61 | |
| F IX | | 0.86 | 0.83 | |
| F X | | 0.82 | 0.70 | |
| F XI | | 0.90 | 0.65 | |
| F XIII | | 0.57 | 0.81 | |

* laboratory was excluded from analysis of between-laboratory variability because of a unique platform, ‡ analysis was limited to routine coagulation parameters, because one laboratory did not assess all tests, ¶ (Platform + Technician), + Laboratories use the platform in different ways (they interpret the manual differently),

Discussions

Variance components that could be attributed to technicians or laboratory procedures were substantial, led to disappointingly low intraclass correlation coefficients for several factors and were pronounced for some of the platforms, probably for those allowing more instrument adjustments. Our results emphasise that variability of parameters between state-of-the-art laboratories remains considerable. This may have a relevant impact on the reproducibility and comparability of the results.

Our results confirm and extend previous reports on a wide variability of coagulation parameters. A wide variability of PT measurements was already noticed with the introduction of early automation coagulometers [13]. An investigation of the World Federation of Haemophilia (WFH) EQA programme and the United Kingdom National External Quality Assessment Scheme (UK NEQAS) conducted in the UK and in emerging countries revealed a good to acceptable variation with regard to PT and aPTT (coefficient of variation [CV]: 10.1-20.4%) but an extensive variation with regard to FVIII, FIX and von Willebrand factor determination (CV 6-154%) [14]. Considerable variations in determination of von Willebrand factor antigen (vWF:Ag) and ristocetin cofactor activity (vWF:RCo) were recognised in investigations of the U.S. College of American Pathologists proficiency testing (CV 3.2-30.9%; n=171 laboratories) [14], the European Concerted Action on Thrombosis and Disabilities Foundation (CV 10-40%; n=181 laboratories) [15] and UK NEQAS (CV 15-50%; n=200) [16].

Agreement with regard to subtherapeutic, therapeutic or supratherapeutic levels in the monitoring of unfractionated heparin using activated partial thromboplastin time or anti-Xa level was very limited in a cross-validation study [17] as well as a study using the results of the annual Ontario Quality Management Program [6]. An interlaboratory agreement of 16-52% and a coefficient of variation of 10.5 – 65% were reported respectively. Another investigation used the results of the Italian External Quality Assessment Scheme and found a wide variability in estimation of quantitative D-Dimers (coefficient of variation up to 47%) [18]. A high degree of variation between laboratories in the identification of coagulation factor inhibitors was shown in several investigations, which used data of external quality assessment programs [7,19-21]. Other investigations identified the type of the reagent [22], the method of the assay [23] and the calibrator [24,25] as factors that bias the measurements between different laboratories and the coagulometer influencing the precision of the measurements [22].

In contrast to these previous investigations, our study illustrates the degree of variation between laboratories using state-of-the-art coagulometers and reagents in relation to the variation of the subjects. We showed the degree of variation for PT, fibrinogen and coagulation factors II, V, VII, VIII, IX, XI and XIII. Furthermore, we characterised factors associated with technicians and laboratory procedures as a relevant source of this variation. These factors include also the choice of the reagents, which may have influenced the higher between-laboratory variability of platform 2 in contrast to platform 3.

Our investigation has several limitations. First, it is an exploratory study with a relative low number of determinations. However, it facilitates future investigations with more determinations and a proper power analysis. Second, “factors associated with technicians and laboratory procedures” are determined as a global factor. The design of

the study did not allow discriminating additional potential factors such as type of the reagents, which are also inadequately addressed by the literature.

Future investigations have to separate these factors to discriminate between organisational factors and single technicians. Third, only fibrinogen determination was analysed for two points in time, because of the study design. If future investigations would determine more parameters two times, better information on differences between parameters would be possible. Forth, only state-of-the-art laboratories specialised in determination of coagulation factors measured the samples in our investigation. The results could possibly be different if routine laboratories would have been investigated.

The strength of our study is that was done it in a nation-wide design including only state-of-the-art coagulation laboratories of all tertiary hospitals. Therefore, a selection bias appears unlikely. Furthermore, a wide range of coagulation parameters was determined. Though, the observed effects were internally confirmed by other parameters. Moreover, it is one of the first investigations that observed the sources of this variation and – to our knowledge – the first investigation, which observed factors associated with technicians and laboratory procedures as possible source for variation of results between laboratories.

The present results suggest that variability of coagulation parameters between specialised laboratories is substantial despite state-of-the-art coagulometers and reagents, great efforts to guarantee precision of laboratory tests and external quality assessment programs. The results indicate that factors associated with technicians and laboratory procedures are a relevant source of this variation. Though, organisational factors are a promising field of work for future investigations on the sources of this variability. Furthermore, efforts to enhance laboratory organisation may be an attractive area of activity for reducing variability of the measurements of coagulation parameters as well as improving quality of laboratory results within laboratories. This hypothesis is supported by the fact that platforms, which allow more adjustments, were associated with a greater variance than others with a more rigid application. Furthermore, parameters, which necessitate a greater effort in structuring operation procedure, have had a lower variance than more simple parameters.

In conclusion, variability between laboratories in determination of coagulation parameters was considerable and variance components that could be attributable to technicians and laboratory procedures were substantial. Once confirmed in larger studies, our findings call for sustained efforts to raise the level of standardization of structures and procedures involved in the quantification of coagulation factors.

References

1. Marshall WJ, Bangert SK, Lapsley M (2012) The clinical utility of laboratory investigations. *Clinical Chemistry*. Philadelphia, PA: Mosby Elsevier. pp. 9.
2. NCCLS (1999) Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline. Wayne, PA: NCCLS.
3. Fritschi J, Raddatz-Muller P, Schmid P, Wuillemin WA (2007) Patient self-management of long-term oral anticoagulation in Switzerland. *Swiss Med Wkly* 137: 252-258.
4. Rouillet S, Pillot J, Freyburger G, Biais M, Quinart A, et al. (2010) Rotation thromboelastometry detects thrombocytopenia and hypofibrinogenaemia during orthotopic liver transplantation. *Br J Anaesth* 104: 422-428.
5. Plebani M, Sanzari MC, Zardo L (2008) Quality control in coagulation testing. *Semin Thromb Hemost* 34: 642-646.
6. Cuker A, Raby A, Moffat KA, Flynn G, Crowther MA (2010) Interlaboratory variation in heparin monitoring: Lessons from the Quality Management Program of Ontario coagulation surveys. *Thromb Haemost* 104: 837-844.
7. Favaloro EJ, Bonar R, Kershaw G, Duncan E, Sioufi J, et al. (2009) Investigations from external quality assurance programs reveal a high degree of variation in the laboratory identification of coagulation factor inhibitors. *Semin Thromb Hemost* 35: 794-805.
8. Meijer P, Haverkate F (2005) External quality assessment and the laboratory diagnosis of thrombophilia. *Semin Thromb Hemost* 31: 59-65.
9. Shahangian S, Labeau KM, Howerton DA (2006) Prothrombin time testing practices: adherence to guidelines and standards. *Clin Chem* 52: 793-794.
10. Gerdsen F, Luxembourg B, Langer F, Bauersachs R, Lindhoff-Last E (2008) A prospective analysis of heparin-platelet factor 4 antibodies in pregnant women treated with the low-molecular-weight heparin, dalteparin. *Blood Coagul Fibrinolysis* 19: 477-481.
11. Banfi G, Del Fabbro M (2009) Biological variation in tests of hemostasis. *Semin Thromb Hemost* 35: 119-126.
12. Schramko A, Suojaranta-Ylinen R, Kuitunen A, Raivio P, Kukkonen S, et al. (2010) Hydroxyethylstarch and gelatin solutions impair blood coagulation after cardiac surgery: a prospective randomized trial. *Br J Anaesth* 104: 691-697.
13. Thomson JM, Taberner DA, Poller L (1990) Automation and prothrombin time: a United Kingdom field study of two widely used coagulometers. *J Clin Pathol* 43: 679-684.
14. Riggio JM, Cooper MK, Leiby BE, Walenga JM, Merli GJ, et al. (2009) Effectiveness of a clinical decision support system to identify heparin induced thrombocytopenia. *J Thromb Thrombolysis* 28: 124-131.
15. Schochl H, Nienaber U, Hofer G, Voelckel W, Jambor C, et al. (2010) Goal-directed coagulation management of major trauma patients using thromboelastometry (ROTEM)-guided administration of fibrinogen concentrate and prothrombin complex concentrate. *Crit Care* 14: R55.
16. Roeloffzen WW, Kluin-Nelemans HC, Mulder AB, Veeger NJ, Bosman L, et al. (2010) In normal controls, both age and gender affect coagulability as measured by thrombelastography. *Anesth Analg* 110: 987-994.
17. Cuker A, Ptashkin B, Konkle BA, Pipe SW, Whinna HC, et al. (2009) Interlaboratory agreement in the monitoring of unfractionated heparin using the anti-factor Xa-correlated activated partial thromboplastin time. *J Thromb Haemost* 7: 80-86.
18. Tripodi A, Chantarangkul V (2007) Performance of quantitative D-dimer methods: results of the Italian external quality assessment scheme. *J Thromb Haemost* 5: 184-185.
19. Warkentin TE (2004) Bivalent direct thrombin inhibitors: hirudin and bivalirudin. *Best Pract Res Clin Haematol* 17: 105-125.

20. Warkentin TE (2004) Randomized trial of warfarin nomograms. *Ann Intern Med* 140: 490-491; author reply 491-492.
21. Greinacher A, Althaus K, Krauel K, Selleng S (2010) Heparin-induced thrombocytopenia. *Hamostaseologie* 30: 17-28.
22. Rauova L, Zhai L, Kowalska MA, Arepally GM, Cines DB, et al. (2006) Role of platelet surface PF4 antigenic complexes in heparin-induced thrombocytopenia pathogenesis: diagnostic and therapeutic implications. *Blood* 107: 2346-2353.
23. Warncke C, Alberio L, Savolainen H, von Allmen R, Baumgartner I, et al. (2008) "HIT on Trousseau" double trouble: acquired coagulopathy with femoral artery thrombosis. *Vasa* 37: 281-284.
24. Arepally GM, Ortel TL (2006) Clinical practice. Heparin-induced thrombocytopenia. *N Engl J Med* 355: 809-817.
25. Warkentin TE, Maurer BT, Aster RH (2007) Heparin-induced thrombocytopenia associated with fondaparinux. *N Engl J Med* 356: 2653-2655; discussion 2653-2655.

CHAPTER 8

Methodological shortcomings and design-related bias in studies investigating diagnostic tests for venous thromboembolic diseases:

Literature survey

Michael Nagler, Patrick M. Bossuyt, Hugo ten Cate, Walter A. Willemin, Livia Faes,
Lucas M. Bachmann

(submitted for publication)

Abstract

Background: The methodological quality of diagnostic accuracy studies for venous thromboembolic diseases and extent of bias due to methodological shortcomings is largely unknown.

Methods and Results: Diagnostic systematic reviews in five target conditions were searched in Medline and Embase. Two-by-two tables and the results of the reviews' quality assessment were extracted for each primary study included in these reviews in duplicate. The median number of patients with and without the target condition and study size was calculated. A meta-analysis, comparing the pooled diagnostic odds ratios of studies unlikely to be biased with studies likely to be biased according to the QUADAS-2 quality items was performed. Eleven systematic reviews summarizing 140 primary studies were included (168 test evaluations; 31'046 patients). Median sample size was 103 (IQR 50-202). The median number of patients with the target condition-needed to calculate a test's sensitivity-was 31 (15-46). The median number of patients without the target condition-needed for the calculation of a test's specificity-was 68 (27-152). Adequate quality assessment tools were used in 5 systematic reviews only (46%). Among the 37 primary studies with QUADAS-2 score available, the risk of bias was estimated to be low for all items in five only (13.5%). Studies unlikely to be biased on all items reported lower DOR (-13.56 (95% CI-46.94 to 19.83); p=0.415).

Conclusions: Diagnostic accuracy in studies on venous thromboembolic diseases are unlikely to report precise point estimates of test performance and are often likely to be biased according to QUADAS-2. Established guidelines to design and report diagnostic tests should be systematically adopted.

Introduction

Little is known about the extent to which established guidelines for the design, evaluation and assessment have been adopted in articles about the diagnosis of venous thromboembolic diseases. Determination of a probable or definitive diagnosis is a crucial step in medical care and the quality of the diagnostic process determines the quality of care to a large amount [1]. It is acknowledged that the use of any diagnostic tests is an intervention that may bring significant harm to patients [2]. Premature application of medical tests in clinical practice may lead to an excess in costs, wrong diagnosis, unnecessary delays in starting appropriate treatment, or even hazardous therapy [3,4]. In addition, it has been recognised that sophisticated and expensive tests that are disseminated without suitable evaluations can subsequently be found to have marginal clinical value and economic benefit. Selection of appropriate diagnostic tests is challenging because there is an accelerating number of tests available [2]. A comprehensive evaluation of diagnostic tests is essential not only for improving quality in health care but limiting costs in an environment of increasing pressure on budgets [5,6]. Diagnostic accuracy studies are key elements of test evaluation and biased results may easily lead to erroneous conclusions on the clinical value of medical tests.

Well known historical examples and some empirical evidence illustrate how methodological shortcomings in diagnostic accuracy studies may lead to wrong conclusions and an unjustified subsequent implementation in clinical practice [7-12]. A number of guidelines and tools for assessing and improving quality of studies evaluating diagnostic tests have been developed in last decades to overcome these problems. In particular, the STARD guideline focuses on the accurate design and reporting of diagnostic accuracy studies and the QUADAS-2 tool assesses the methodological quality of studies to be used in systematic reviews and meta-analyses [2,13,14]. However, to what extent these guidelines have been adopted in clinical practice is largely unknown. Furthermore, it is not known to what extent methodological shortcomings lead to bias in diagnostic accuracy estimates. Venous thromboembolic diseases contribute to the health burden and large resources are necessary for diagnosis and treatment of these patients [15,16]. Uncertainties in the diagnostic process of patients with venous thromboembolism may contribute to adverse outcomes and excess in costs.

In this study we systematically assessed the methodological quality of studies evaluating diagnostic tests for venous thromboembolic diseases and quantified the direction and impact of biased studies on average diagnostic performance.

Methods

Search strategy, data sources and inclusion criteria

We performed an electronic literature search to identify systematic reviews and meta-analyses as well as subsequent original studies on diagnostic accuracy of test used for venous thromboembolic diseases. To get a comprehensive dataset, we included all disease entities in the search strategy that are associated with venous thromboembolism: venous thrombosis, pulmonary embolism, lower extremity deep vein thrombosis, heparin-induced thrombocytopenia and disseminated intravascular coagulation. To identify diagnostic accuracy studies, a validated strategy was used [17]. The search strategy was developed by two authors (LMN, MN) and tested on a set of 10 index publications. Details are given in the appendix. MEDLINE, and EMBASE were searched from inception to April 2014. The database search was complemented by screening reference lists of included studies.

Inclusion criteria were (1) systematic reviews of diagnostic accuracy studies, (2) index test used to identify one of the mentioned disease entities, and (4) sufficient details to generate a 2 by 2 contingency table. To be included in meta-analysis determining the effect of design-related bias, original studies needed to have a QUADAS-2 guided assessment of methodological quality by previous investigators of the respective diagnostic test. No language restrictions were applied.

Literature search and selection of publications for full-text review was done by one author (MN). Full-text review and assessment for eligibility was independently done by two authors (MN, LF). A third reviewer (LMB) arbitrated unclear cases. This was necessary for 13 extracted items.

Data extraction

All data were extracted in a standardized extraction form. The following data were drawn out of the included systematic reviews: number of studies, employment and type of a quality assessment tool, and QUADAS (Version 1 and 2) results assigned to individual studies (if applied). Primary studies were analyzed to extract: sample size, numbers of affected and unaffected patients, prevalence, and numbers of the 2 by 2 contingency table (true positives, false positives, false negatives, true negatives). Two reviewers (LF, LMB) independently did data extraction and disagreement was resolved by consensus.

Measures of methodological quality

Measures of different dimensions were used to determine methodological quality. First, included systematic reviews were assessed with regard to adequate application of quality assessment tools. Second, primary studies were analysed with regard to adequate sample sizes [18]. And third, methodological quality of primary studies as assessed by previous investigators was recorded, expressed as QUADAS-2 result. We de-

cided to rely on the assessment of previous authors for two reasons: (1) application of QUADAS-2 has to be done in the context of and adapted to the respective research question, what is better done in a systematic review focussing on a specific diagnostic test, and (2) to avoid possible bias due to QUADAS-2 scoring of investigators studying methodological quality.

Statistical analyses

From extracted two-by-two tables we calculated the diagnostic odds ratio (DOR) per study. DOR was then pooled per QUADAS item (bias unlikely vs. likely or missing information). In a meta-regression analysis, we investigated the effect of high study quality (bias unlikely in all QUADAS items) on the pooled DOR.

Statistical analyses were performed using the Stata 11.2 statistical software package using the *metan* and *metareg* commands. (StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.)

Results

Study selection

Literature search identified 833 articles, including 68 duplicates (**Figure 1**). It was last updated in March 2014. Based on the inclusion criteria, we selected 75 records for full-text screening and rejected 690. Reviewing the full-text articles, we excluded 23 because of the publication type (no systematic review), 40 because of insufficient data, and one article because of a different scope. Finally, we included 11 systematic reviews [19-29] and 140 primary studies subsequently [30-168]. Multiple evaluations were done in some studies resulting in 168 test evaluations and 31'046 patients. For 36 primary studies [34,37,42,49,50,53,54,69,70,75-79,83,88,90,91,93,94,98-101,108,110, 123,126, 132,136,138,139,150,160,161,167] QUADAS-2 scores were available in 3 systematic [21,27,28]. These studies were included in meta-analysis of design-related bias.

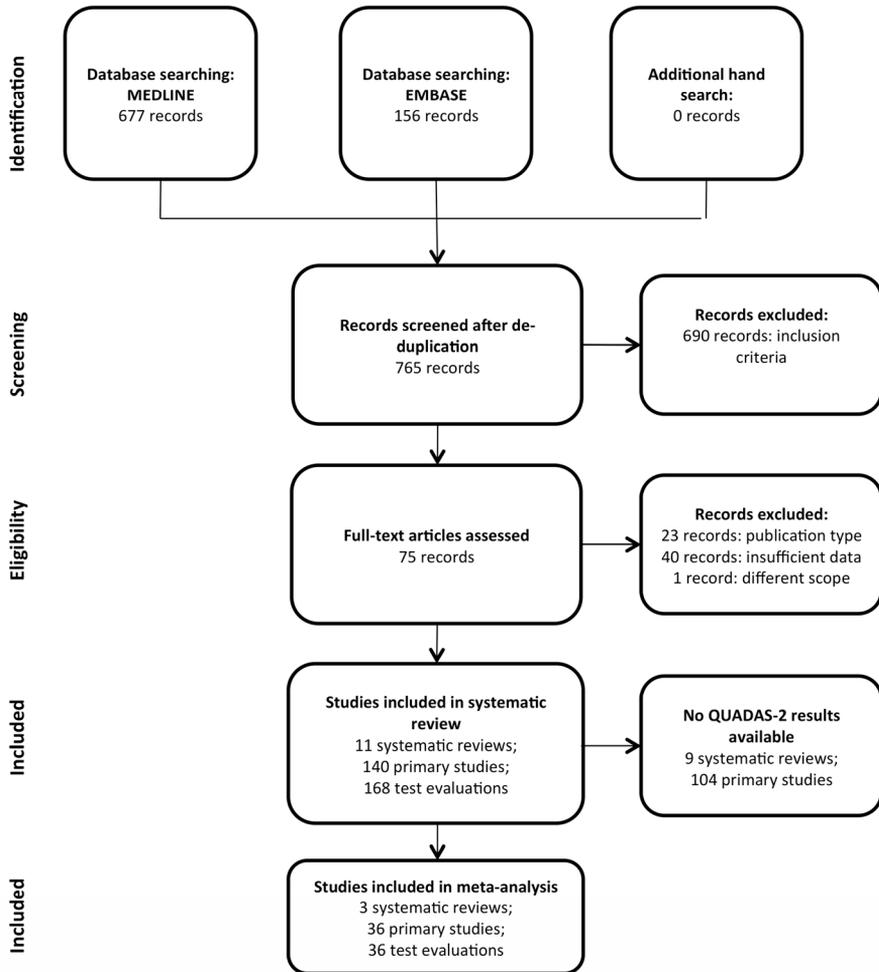


Figure 1: PRISMA flow chart

Study characteristics and sample size

The characteristics of included diagnostic accuracy studies are displayed in **Table 1**. Studies covered the whole spectrum of diagnostic problems associated with venous thromboembolism and a wide range of diagnostic tests. The frequency distribution of diseases and tests points towards a representative sample. Systematic reviews included a median of 10 primary studies with a inter-quartile range (IQR) of 8 to 15 (range 7 [24] to 67 [26]). Median prevalence of the target condition in primary studies was 31.4% (IQR 22.6-42.9%). The studies included a median number of 31 patients with the disease (IQR 15-46), as well as 68 without the disease (IQR 27-152). The median sample size of primary diagnostic accuracy studies was 103, with an IQR of 50 to 202.

Table 1: Summary characteristics of studies included in systematic review

| Characteristics | Results |
|---|---------------------|
| Systematic reviews (<i>number</i>) | 11 |
| Number of included studies – (<i>median;IQR</i>) | 10 (8-15) |
| Primary studies (<i>number</i>) | 140 |
| Test evaluations (number) | 168 |
| Sample size (median; IQR) | 103 (50–202) |
| Prevalence (median; IQR) | 30.7% (22.3%-42.9%) |
| Patients with the target condition (<i>median; IQR</i>) | 31 (15-46) |
| Patients without the target condition (<i>median; IQR</i>) | 68 (27-152) |
| Diseases | |
| Deep vein thrombosis | 25 (14.9%) |
| Pulmonary embolism | 111 (66.1%) |
| Upper extremity deep vein thrombosis | 19 (11.3%) |
| Heparin-induced thrombocytopenia | 13 (7.7%) |
| Diagnostic tests | |
| Compression ultrasonography | 24 (14.3%) |
| Doppler ultrasonography | 10 (6.0%) |
| Duplex ultrasonography | 2 (1.2%) |
| Computed tomography | 21 (12.5%) |
| D-Dimer | 42 (25.0%) |
| Brain natriuretic peptide/ N-terminal pro-brain natriuretic peptide | 8 (4.8%) |
| Ventilation perfusion lung scan | 10 (6.0%) |
| Magnetic resonance angiography | 5 (3.0%) |
| Echocardiography | 2 (1.2%) |
| Lung ultrasonography | 10 (6.0%) |
| Capnography | 14 (8.3%) |
| 4Ts score | 13 (7.7%) |
| Rheography | 3 (1.8%) |
| Others | 4 (2.4%) |
| Quality assessments | |
| Studies with low risk of bias (<i>number; percent</i>) | 5 (13.5%) |

IQR, inter-quartile range

Quality assessment

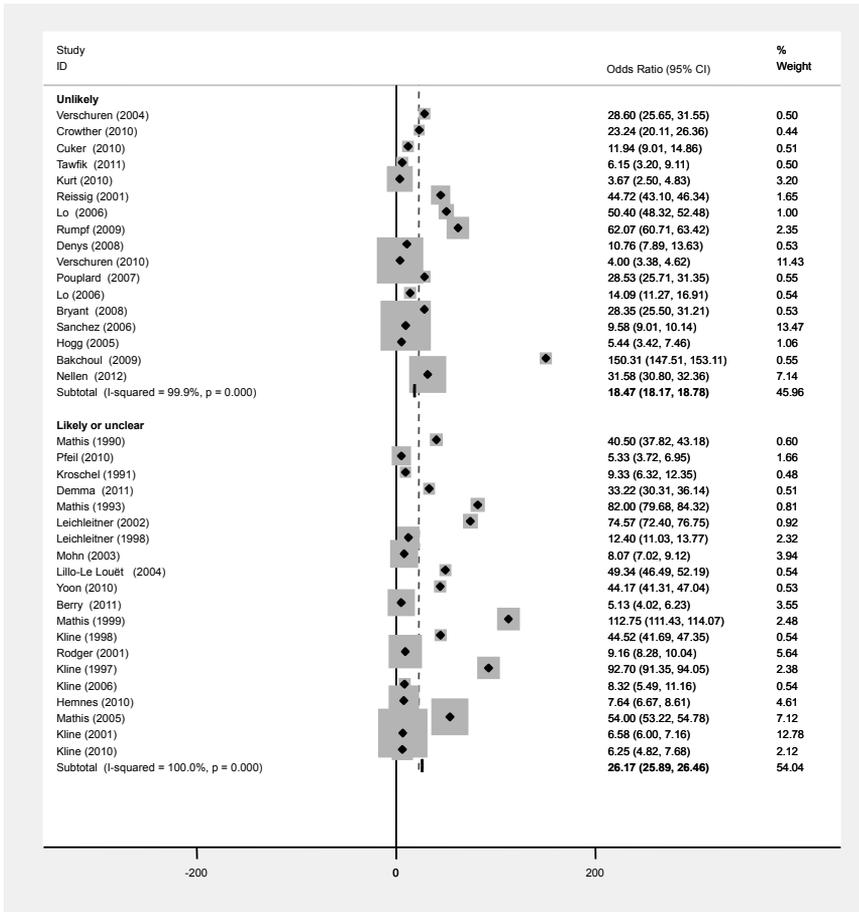
All included systematic reviews assessed quality of included diagnostic accuracy studies, but only 5 (45.5%) used the validated and recommended QUADAS/QUADAS-2 tool [14,169]. Six reviews used a self-constructed questionnaire (54.5%) and four of them used it even though QUADAS/QUADAS-2 was already available at the date of publication. Among the studies with a score of the QUADAS-2 tool available, the risk of bias was estimated to be low for all items in five out of 37 only (13.5%). In particular, the risk of bias was estimated to be “high” or “unclear” with regard to the reference test (74.2%), the index test (64.5%), and the patient selection (64.5%). No quality assessment was done in two reviews.

Extent of bias

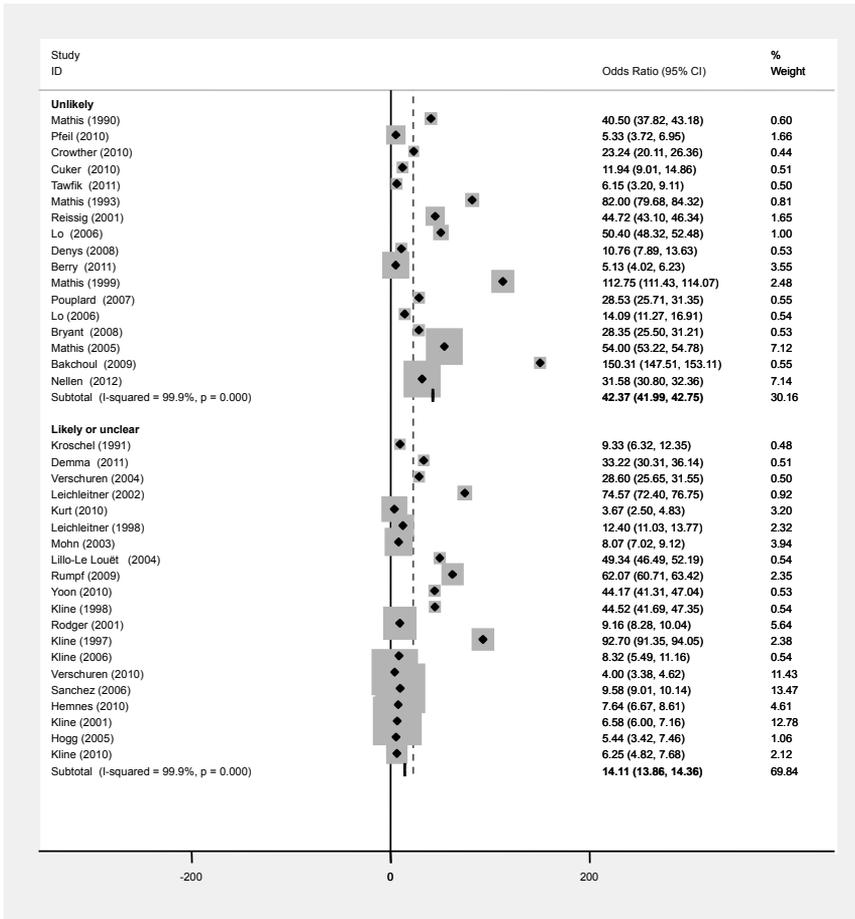
In **Figure 2**, the diagnostic odds ratios are shown after stratifying the studies according to the estimated risk of bias. Results are shown for every item of the QUADAS-2 tool (patient selection, index test, reference standard, and flow and timing). As mentioned above, in the majority of studies the risk of bias was rated to be „likely“ or „unclear“. Between studies with and without a fulfilled quality criterion, the diagnostic odds ratio varied to a large amount. Interestingly, the degree and the direction of bias differed substantially between QUADAS items. For “patient selection” and “reference test” the diagnostic odds ratio is nearly as double as large in studies without fulfilment (for details see Figure 2). In contrast, for “index test” and “flow & timing”, the diagnostic odds ratio was three times as high in studies with fulfilled criteria than in studies without. Studies unlikely to be biased on all items reported lower DOR (-13.56 (95% CI-46.94 to 19.83); p=0.415).

Figure 2: Risk of bias due to methodological shortcomings in diagnostic accuracy studies for venous thrombo-embolic diseases. Diagnostic odds ratios are shown for studies with (“risk of bias unlikely”) and without (“risk of bias likely or unclear”) fulfilment of the individual item of the QUADAS-2 quality assessment tool.

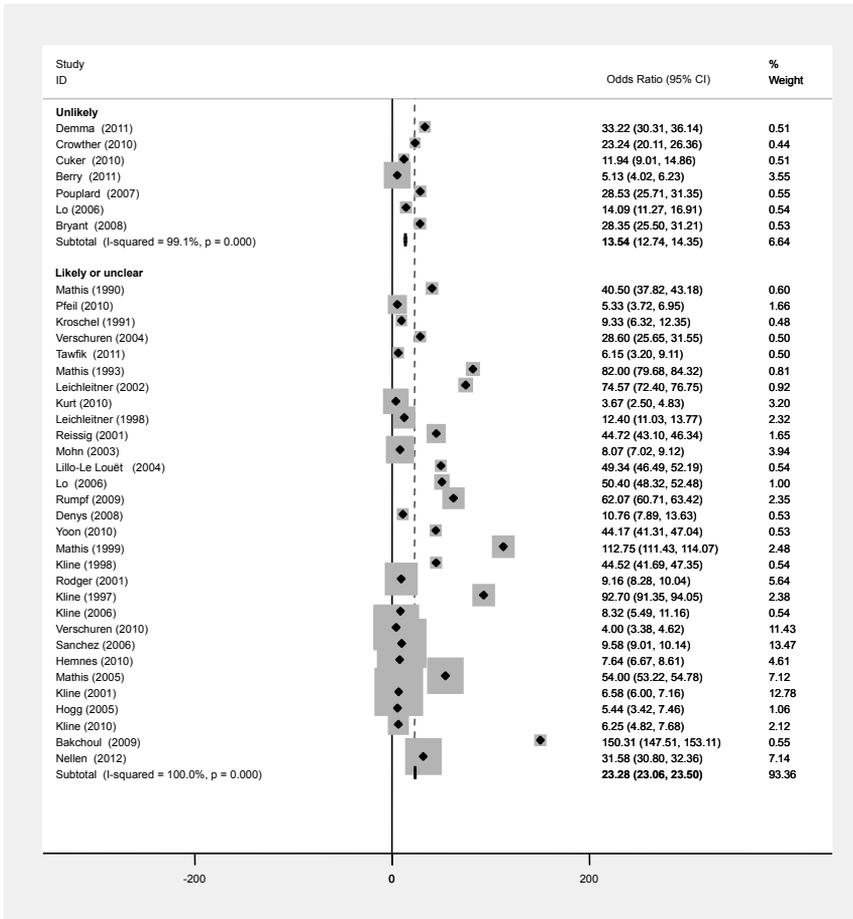
Risk of selection bias



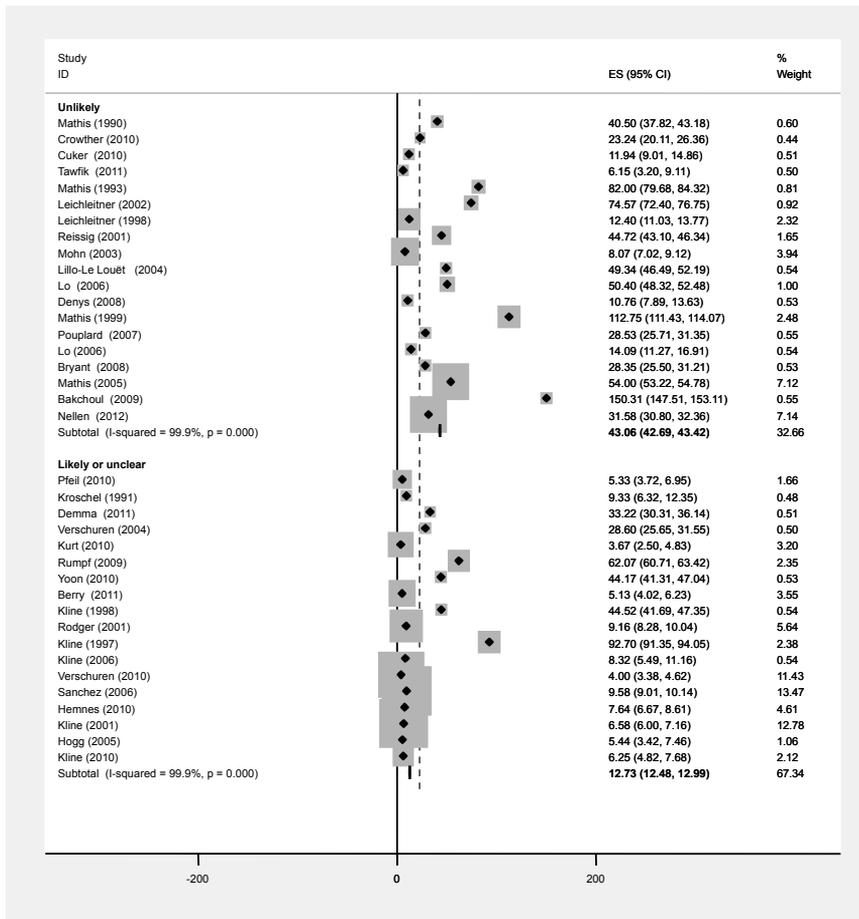
Risk of bias index



Risk of bias due to reference test



Risk of bias due to flow & timing



Discussion

By assessing a positive selection of studies evaluating diagnostic tests for venous thromboembolic diseases we found major shortcomings in the methodological quality and empirical evidence for design-related biases in diagnostic accuracy measures. The sample size of the diagnostic accuracy studies was even more limited than previous recorded in a selection of high-ranked general and specialised scientific journals [18]. Furthermore, application of appropriate quality assessment tools in systematic reviews and adequate presentation of the results was done in the minority of the studies only. In addition, the risk of bias was estimated to be high or unclear in the majority of stud-

ies. Notably, a major systematic bias of the diagnostic odds ratio was assessed in studies with non-fulfilment of quality criteria specified by the QUADAS-2 tool.

We are unaware of previous studies investigating extent and effects of methodological shortcomings systematically in diagnostic accuracy studies for venous thromboembolic diseases. Bachmann et al. studied sample sizes and a-priori sample size calculations in a selection of high-ranked general and specialised scientific journals [18]. Sample sizes we recorded in diagnostic accuracy studies for venous thromboembolism were even more limited. Our results are in-line with previous publications in general medical journals confirming that methodological shortcomings are common and the quality of reporting restricted [7,9]. Some evidence exists regarding a systematic bias due to methodological shortcomings. Already in 1999 Lijmer and colleagues reported on systematic overestimation of the diagnostic ability of a test if particular methodological requirements are not met [7]. Interestingly, the issues they found to have a high risk correspond very well to the domains of the QUADAS-2 tool we identified: “patient selection” and “reference test”. Rutjes et al also focussed on various methodological factors associated with a bias in estimated diagnostic accuracy [9]. In agreement with our results, authors identified issues associated with patient selection as particularly sensitive to bias. This observation is further supported by a systematic review and meta-analysis focussing on magnetic resonance imaging for detecting silicone breast implant ruptures [10]. The authors identified patient selection procedure as particularly sensitive to bias. Other authors mentioned disease prevalence and details of data analysis as potential sources of bias [170]. In contrast to these previous investigations, we observed a relationship between bias and quality status according to QUADAS-2.

The strength of our investigation is that we studied methodological issues in a large number of studies investigating a broad range of diagnostic tests for diagnosing all important diseases associated with venous thromboembolism. Arguably our sample is a positive selection of the full body of evidence, since only studies included in systematic reviews were considered. Thus, the problem might be worse when considering the complete diagnostic literature on venous thromboembolic diseases. Furthermore, we assessed extent of bias regarding methodological deficiencies focussing on the most recent and recommended quality assessment tool QUADAS-2. In contrast to previous investigations, we were able to determine the direction and to quantify extent of bias for individual QUADAS-2 domains.

Our results suggest that diagnostic accuracy in studies observing test for venous thromboembolic diseases are biased by methodological shortcomings. This observation call for increased efforts to implement current guidelines for reporting and assessment of methodological quality [2,14]. Several authors demanded a phased evaluation, according to the evaluation required for FDA-approval of new drugs [3,171-173]. In a first phase, the analytical characteristics and the technical accuracy including reproducibility

shall be determined. In a second phase, the diagnostic accuracy determined by an adequate study design will be investigated. These studies will be complemented by a phase three determining health outcomes (mortality and morbidity) of using the test. Afterwards, the cost effectiveness must be determined, decision-making algorithms developed and the organisational impact evaluated. The advantage of a phased evaluation is that further evaluation will be stopped after insufficient results at an early stage. Thus, significant harm to patients associated with a premature implementation of medical tests will be avoided. Furthermore, a relevant amount of costs that are associated with the use of tests with unclear value can be saved.

In conclusion, studies investigating diagnostic accuracy of tests used to identify venous thromboembolic diseases are flawed by methodological shortcomings. They are unlikely to report precise point estimates of test performance and are often likely to be biased according to QUADAS-2. Biased studies tended to exaggerate test performance. Established guidelines to design and report diagnostic tests should be systematically adopted in the diagnostic literature on venous thromboembolic diseases.

References

1. Knottnerus JA (2009) *The Evidence Base of Clinical Diagnosis*. Chichester, West Sussex, UK: Wiley-Blackwell. i-xiii p.
2. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy*. *Clin Chem* 49: 1-6.
3. Lijmer JG, Leeflang M, Bossuyt PM (2009) Proposals for a phased evaluation of medical tests. *Med Decis Making* 29: E13-21.
4. Epner PL, Gans JE, Graber ML (2013) When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf* 22 Suppl 2: ii6-ii10.
5. Bossuyt PM, Lijmer JG, Mol BW (2000) Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 356: 1844-1847.
6. Graber ML (2013) The incidence of diagnostic error in medicine. *BMJ Qual Saf* 22 Suppl 2: ii21-ii27.
7. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, et al. (1999) Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282: 1061-1066.
8. Nierenberg AA, Feinstein AR (1988) How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 259: 1699-1702.
9. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, et al. (2006) Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 174: 469-476.
10. Song JW, Kim HM, Bellfi LT, Chung KC (2011) The effect of study design biases on the diagnostic accuracy of magnetic resonance imaging for detecting silicone breast implant ruptures: a meta-analysis. *Plast Reconstr Surg* 127: 1029-1044.
11. Fletcher RH (1986) Carcinoembryonic antigen. *Ann Intern Med* 104: 66-73.
12. Lensing AW, Hirsh J (1993) 125I-fibrinogen leg scanning: reassessment of its role for the diagnosis of venous thrombosis in post-operative patients. *Thromb Haemost* 69: 2-7.
13. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3: 25.
14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155: 529-536.
15. Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, et al. (1998) Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med* 158: 585-593.
16. Spencer FA, Gurwitz JH, Schulman S, Linkins LA, Crowther MA, et al. (2014) Venous thromboembolism in older adults: A community-based study. *Am J Med* 127: 530-537 e533.
17. Bachmann LM, Coray R, Estermann P, Ter Riet G (2002) Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc* 9: 653-658.
18. Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM (2006) Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 332: 1127-1129.
19. Wells PS, Owen C, Doucette S, Fergusson D, Tran H (2006) Does this patient have deep vein thrombosis? *JAMA* 295: 199-207.
20. Di Nisio M, Van Sluis GL, Bossuyt PM, Buller HR, Porreca E, et al. (2010) Accuracy of diagnostic tests for clinically suspected upper extremity deep vein thrombosis: a systematic review. *J Thromb Haemost* 8: 684-692.
21. Cuker A, Gimotty PA, Crowther MA, Warkentin TE (2012) Predictive value of the 4Ts scoring system for heparin-induced thrombocytopenia: a systematic review and meta-analysis. *Blood* 120: 4160-4167.

22. Pomero F, Dentali F, Borretta V, Bonzini M, Melchio R, et al. (2013) Accuracy of emergency physician-performed ultrasonography in the diagnosis of deep-vein thrombosis: a systematic review and meta-analysis. *Thromb Haemost* 109: 137-145.
23. Harvey RT, Gefter WB, Hrunig JM, Langlotz CP (2000) Accuracy of CT angiography versus pulmonary angiography in the diagnosis of acute pulmonary embolism: evaluation of the literature with summary ROC curve analysis. *Acad Radiol* 7: 786-797.
24. Eng J, Krishnan JA, Segal JB, Bolger DT, Tamariz LJ, et al. (2004) Accuracy of CT in the diagnosis of pulmonary embolism: a systematic literature review. *AJR Am J Roentgenol* 183: 1819-1827.
25. Hayashino Y, Goto M, Noguchi Y, Fukui T (2005) Ventilation-perfusion scanning and helical CT in suspected pulmonary embolism: meta-analysis of diagnostic performance. *Radiology* 234: 740-748.
26. Roy PM, Colombet I, Durieux P, Chatellier G, Sors H, et al. (2005) Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. *BMJ* 331: 259.
27. Manara A, D'Hoore W, Thys F (2013) Capnography as a diagnostic tool for pulmonary embolism: a meta-analysis. *Ann Emerg Med* 62: 584-591.
28. Squizzato A, Rancan E, Dentali F, Bonzini M, Guasti L, et al. (2013) Diagnostic accuracy of lung ultrasound for pulmonary embolism: a systematic review and meta-analysis. *J Thromb Haemost* 11: 1269-1278.
29. Cavallazzi R, Nair A, Vasu T, Marik PE (2008) Natriuretic peptides in acute pulmonary embolism: a systematic review. *Intensive Care Med* 34: 2147-2156.
30. Aguilar C, Martinez A, Martinez A, Del Rio C, Vazquez M, et al. (2002) Diagnostic value of d-dimer in patients with a moderate pretest probability of deep venous thrombosis. *Br J Haematol* 118: 275-277.
31. Anderson DR, Kovacs MJ, Kovacs G, Stiell I, Mitchell M, et al. (2003) 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). *J Thromb Haemost* 1: 645-651.
32. Anderson DR, Wells PS, Stiell I, MacLeod B, Simms M, et al. (2000) Management of patients with suspected deep vein thrombosis in the emergency department: combining use of a clinical diagnosis model with D-dimer testing. *J Emerg Med* 19: 225-230.
33. Baarslag HJ, van Beek EJ, Koopman MM, Reekers JA (2002) Prospective study of color duplex ultrasonography compared with contrast venography in patients suspected of having deep venous thrombosis of the upper extremities. *Ann Intern Med* 136: 865-872.
34. Bakchoul T, Giptner A, Najaoui A, Bein G, Santoso S, et al. (2009) Prospective evaluation of PF4/heparin immunoassays for the diagnosis of heparin-induced thrombocytopenia. *J Thromb Haemost* 7: 1260-1265.
35. Bates SM, Kearon C, Crowther M, Linkins L, O'Donnell M, et al. (2003) A diagnostic strategy involving a quantitative latex D-dimer assay reliably excludes deep venous thrombosis. *Ann Intern Med* 138: 787-794.
36. Baxter GM, Kincaid W, Jeffrey RF, Millar GM, Porteous C, et al. (1991) Comparison of colour Doppler ultrasound with venography in the diagnosis of axillary and subclavian vein thrombosis. *Br J Radiol* 64: 777-781.
37. Berry C, Tcherniantchouk O, Ley EJ, Salim A, Mirocha J, et al. (2011) Overdiagnosis of heparin-induced thrombocytopenia in surgical ICU patients. *J Am Coll Surg* 213: 10-17; discussion 17-18.
38. Blaivas M, Lambert MJ, Harwood RA, Wood JP, Konicki J (2000) Lower-extremity Doppler for deep venous thrombosis--can emergency physicians be accurate and fast? *Acad Emerg Med* 7: 120-126.
39. Blum AG, Delfau F, Grignon B, Beurrier D, Chabot F, et al. (1994) Spiral-computed tomography versus pulmonary angiography in the diagnosis of acute massive pulmonary embolism. *Am J Cardiol* 74: 96-98.
40. Bounameaux H, Schneider PA, Slosman D, de Moerloose P, Reber G (1990) Plasma D-dimer in suspected pulmonary embolism: a comparison with pulmonary angiography and ventilation--perfusion scintigraphy. *Blood Coagul Fibrinolysis* 1: 577-579.
41. Bova C, Greco F, Misuraca G, Serafini O, Crocco F, et al. (2003) Diagnostic utility of echocardiography in patients with suspected pulmonary embolism. *Am J Emerg Med* 21: 180-183.

42. Bryant A, Low J, Austin S, Joseph JE (2008) Timely diagnosis and management of heparin-induced thrombocytopenia in a frequent request, low incidence single centre using clinical 4T's score and particle gel immunoassay. *Br J Haematol* 143: 721-726.
43. Bucek RA, Koca N, Reiter M, Haumer M, Zontsich T, et al. (2002) Algorithms for the diagnosis of deep-vein thrombosis in patients with low clinical pretest probability. *Thromb Res* 105: 43-47.
44. Chance JF, Abbitt PL, Tegtmeier CJ, Powers RD (1991) Real-time ultrasound for the detection of deep venous thrombosis. *Annals of Emergency Medicine* 20: 494-496.
45. Christiansen F (1997) Diagnostic imaging of acute pulmonary embolism. *Acta Radiol Suppl* 410: 1-33.
46. Christiansen F, Kellerth T, Andersson T, Ragnarsson A, Hjortevang F (1996) Ultrasound at scintigraphic "intermediate probability of pulmonary embolism". *Acta Radiol* 37: 14-17.
47. Constans J, Salmi LR, Sevestre-Pietri MA, Perusat S, Nguon M, et al. (2008) A clinical prediction score for upper extremity deep venous thrombosis. *Thromb Haemost* 99: 202-207.
48. Crisp JG, Lovato LM, Jang TB (2010) Compression ultrasonography of the lower extremity with portable vascular ultrasonography can accurately detect deep venous thrombosis in the emergency department. *Annals of Emergency Medicine* 56: 601-610.
49. Crowther MA, Cook DJ, Albert M, Williamson D, Meade M, et al. (2010) The 4Ts scoring system for heparin-induced thrombocytopenia in medical-surgical intensive care unit patients. *J Crit Care* 25: 287-293.
50. Cuker A, Arepally G, Crowther MA, Rice L, Datko F, et al. (2010) The HIT Expert Probability (HEP) Score: a novel pre-test probability model for heparin-induced thrombocytopenia based on broad expert opinion. *J Thromb Haemost* 8: 2642-2650.
51. de Groot MR, van Marwijk Kooy M, Pouwels JG, Engelage AH, Kuipers BF, et al. (1999) The use of a rapid D-dimer blood test in the diagnostic work-up for pulmonary embolism: a management study. *Thromb Haemost* 82: 1588-1592.
52. Dean AJ, Davies S (2004) Accuracy of Resident-performed Limited Compression Ultrasonography for the Diagnosis of Lower-extremity Deep Vein Thrombosis. *Acad Emerg Med (Supplement SAEM Annual Meeting Abstracts)* 11: 584.
53. Demma LJ, Winkler AM, Levy JH (2011) A diagnosis of heparin-induced thrombocytopenia with combined clinical and laboratory methods in cardiothoracic surgical intensive care unit patients. *Anesth Analg* 113: 697-702.
54. Denys B, Stove V, Philippe J, Devreese K (2008) A clinical-laboratory approach contributing to a rapid and reliable diagnosis of heparin-induced thrombocytopenia. *Thromb Res* 123: 137-145.
55. Drucker EA, Rivitz SM, Shepard JA, Boiselle PM, Trotman-Dickenson B, et al. (1998) Acute pulmonary embolism: assessment of helical CT for diagnosis. *Radiology* 209: 235-241.
56. Falk RL, Smith DF (1987) Thrombosis of upper extremity thoracic inlet veins: diagnosis with duplex Doppler sonography. *AJR Am J Roentgenol* 149: 677-682.
57. Farahmand S, Farnia M, Shahriaran S, Khashayar P (2011) The accuracy of limited B-mode compression technique in diagnosing deep venous thrombosis in lower extremities. *Am J Emerg Med* 29: 687-690.
58. Ferretti GR, Bosson JL, Buffaz PD, Ayanian D, Pison C, et al. (1997) Acute pulmonary embolism: role of helical CT in 164 patients with intermediate probability at ventilation-perfusion scintigraphy and normal results at duplex US of the legs. *Radiology* 205: 453-458.
59. Fojtik PJ, Handly N, Constantino TG (2003) Accuracy of Emergency Medicine Performed Compression Ultrasound for the Evaluation of Deep Venous Thrombosis. *Acad Emerg Med (Supplement SAEM Annual Meeting Abstracts)* 10: 426.
60. Frazee BW, Snoey ER, Levitt A (2001) Emergency Department compression ultrasound to diagnose proximal deep vein thrombosis. *J Emerg Med* 20: 107-112.
61. Garg K, Welsh CH, Feyerabend AJ, Subber SW, Russ PD, et al. (1998) Pulmonary embolism: diagnosis with spiral CT and ventilation-perfusion scanning--correlation with pulmonary angiographic results or clinical outcome. *Radiology* 208: 201-208.

62. Goldhaber SZ, Simons GR, Elliott CG, Haire WD, Toltzis R, et al. (1993) Quantitative plasma D-dimer levels among patients undergoing pulmonary angiography for suspected pulmonary embolism. *JAMA* 270: 2819-2822.
63. Goodman LR, Curtin JJ, Mewissen MW (1995) Detection of pulmonary embolism in patients with unresolved clinical and scintigraphic diagnosis: helical CT versus angiography. *AJR Am J Roentgenol*: 1369-1374.
64. Grist TM, Sostman HD, MacFall JR, Foo TK, Spritzer CE, et al. (1993) Pulmonary angiography with MR imaging: preliminary clinical experience. *Radiology* 189: 523-530.
65. Gupta A, Frazer CK, Ferguson JM, Kumar AB, Davis SJ, et al. (1999) Acute pulmonary embolism: diagnosis with MR angiography. *Radiology* 210: 353-359.
66. Haire WD, Lynch TG, Lund GB, Lieberman RP, Edney JA (1991) Limitations of magnetic resonance imaging and ultrasound-directed (duplex) scanning in the diagnosis of subclavian vein thrombosis. *J Vasc Surg* 13: 391-397.
67. Heit JA, Meyers BJ, Plumhoff EA, Larson DR, Nichols WL (2000) Operating characteristics of automated latex immunoassay fibrin D-dimer tests in the diagnosis of angiographically-defined acute pulmonary embolism. *Thromb Haemost* 83: 970.
68. Heit JA, Minor TA, Andrews JC, Larson DR, Li H, et al. (1999) Determinants of plasma fibrin D-dimer sensitivity for acute pulmonary embolism as defined by pulmonary angiography. *Arch Pathol Lab Med* 123: 235-240.
69. Hemnes AR, Newman AL, Rosenbaum B, Barrett TW, Zhou C, et al. (2010) Bedside end-tidal CO₂ tension as a screening tool to exclude pulmonary embolism. *Eur Respir J* 35: 735-741.
70. Hogg K, Dawson D, Tabor T, Tabor B, Mackway-Jones K (2005) Respiratory dead space measurement in the investigation of pulmonary embolism in outpatients with pleuritic chest pain. *Chest* 128: 2195-2202.
71. Hull RD, Raskob GE (1991) Low-probability lung scan findings: a need for change. *Ann Intern Med* 114: 142-143.
72. Jacoby J, Cesta M, Axelband J, Melanson S, Heller M, et al. (2007) Can emergency medicine residents detect acute deep venous thrombosis with a limited, two-site ultrasound examination? *J Emerg Med* 32: 197-200.
73. Jang T, Docherty M, Aubin C, Polites G (2004) Resident-performed compression ultrasonography for the detection of proximal deep vein thrombosis: fast and accurate. *Acad Emerg Med* 11: 319-322.
74. Kearon C, Ginsberg JS, Douketis J, Crowther M, Brill-Edwards P, et al. (2001) Management of suspected deep venous thrombosis in outpatients by using clinical assessment and D-dimer testing. *Ann Intern Med* 135: 108-111.
75. Kline JA, Arunachlam M (1998) Preliminary study of the capnogram waveform area to screen for pulmonary embolism. *Ann Emerg Med* 32: 289-296.
76. Kline JA, Hogg M (2006) Measurement of expired carbon dioxide, oxygen and volume in conjunction with pretest probability estimation as a method to diagnose and exclude pulmonary venous thromboembolism. *Clin Physiol Funct Imaging* 26: 212-219.
77. Kline JA, Hogg MM, Courtney DM, Miller CD, Jones AE, et al. (2010) D-dimer and exhaled CO₂/O₂ to detect segmental pulmonary embolism in moderate-risk patients. *Am J Respir Crit Care Med* 182: 669-675.
78. Kline JA, Israel EG, Michelson EA, O'Neil BJ, Plewa MC, et al. (2001) Diagnostic accuracy of a bedside D-dimer assay and alveolar dead-space measurement for rapid exclusion of pulmonary embolism: a multicenter study. *JAMA* 285: 761-768.
79. Kline JA, Meek S, Boudrow D, Warner D, Colucciello S (1997) Use of the alveolar dead space fraction (V_d/V_t) and plasma D-dimers to exclude acute pulmonary embolism in ambulatory patients. *Acad Emerg Med* 4: 856-863.

80. Kline JA, O'Malley PM, Tayal VS, Snead GR, Mitchell AM (2008) Emergency clinician-performed compression ultrasonography for deep venous thrombosis of the lower extremity. *Ann Emerg Med* 52: 437-445.
81. Knudson GJ, Wiedmeyer DA, Erickson SJ, Foley WD, Lawson TL, et al. (1990) Color Doppler sonographic imaging in the assessment of upper-extremity deep venous thrombosis. *AJR Am J Roentgenol* 154: 399-403.
82. Kraaijenhagen RA, Piovella F, Bernardi E, Verlato F, Beckers EA, et al. (2002) Simplification of the diagnostic management of suspected deep vein thrombosis. *Arch Intern Med* 162: 907-911.
83. Kroschel U, Seitz K, Reuss J, Rettenmaier G (1991) [Sonographic imaging of lung emboli. Results of a prospective study]. *Ultraschall Med* 12: 263-268.
84. Kruger S, Graf J, Merx MW, Koch KC, Kunz D, et al. (2004) Brain natriuretic peptide predicts right heart failure in patients with acute pulmonary embolism. *Am Heart J* 147: 60-65.
85. Kruip MJ, Slob MJ, Schijen JH, van der Heul C, Buller HR (2002) Use of a clinical decision rule in combination with D-dimer concentration in diagnostic workup of patients with suspected pulmonary embolism: a prospective management study. *Arch Intern Med* 162: 1631-1635.
86. Kruit WH, de Boer AC, Sing AK, van Roon F (1991) The significance of venography in the management of patients with clinically suspected pulmonary embolism. *J Intern Med* 230: 333-339.
87. Kucher N, Printzen G, Doernhoefer T, Windecker S, Meier B, et al. (2003) Low pro-brain natriuretic peptide levels predict benign clinical outcome in acute pulmonary embolism. *Circulation* 107: 1576-1578.
88. Kurt OK, Alpar S, Sipit T, Guven SF, Erturk H, et al. (2010) The diagnostic role of capnography in pulmonary embolism. *Am J Emerg Med* 28: 460-465.
89. Kutinsky I, Blakley S, Roche V (1999) Normal D-dimer levels in patients with pulmonary embolism. *Arch Intern Med* 159: 1569-1572.
90. Lechleitner P, Raneburger W, Gamper G, Riedl B, Benedikt E, et al. (1998) Lung sonographic findings in patients with suspected pulmonary embolism. *Ultraschall Med* 19: 78-82.
91. Lechleitner P, Riedl B, Raneburger W, Gamper G, Theurl A, et al. (2002) Chest sonography in the diagnosis of pulmonary embolism: a comparison with MRI angiography and ventilation perfusion scintigraphy. *Ultraschall Med* 23: 373-378.
92. Leclercq MG, Lutisan JG, van Marwijk Kooy M, Kuipers BF, Oostdijk AH, et al. (2003) Ruling out clinically suspected pulmonary embolism by assessment of clinical probability and D-dimer levels: a management study. *Thromb Haemost* 89: 97-103.
93. Lillo-Le Louet A, Boutouyrie P, Alhenc-Gelas M, Le Beller C, Gautier I, et al. (2004) Diagnostic score for heparin-induced thrombocytopenia after cardiopulmonary bypass. *J Thromb Haemost* 2: 1882-1888.
94. Lo GK, Juhl D, Warkentin TE, Sigouin CS, Eichler P, et al. (2006) Evaluation of pretest clinical score (4 T's) for the diagnosis of heparin-induced thrombocytopenia in two clinical settings. *J Thromb Haemost* 4: 759-765.
95. Logeart D, Lecuyer L, Thabut G, Tabet JY, Tartiere JM, et al. (2007) Biomarker-based strategy for screening right ventricular dysfunction in patients with non-massive pulmonary embolism. *Intensive Care Med* 33: 286-292.
96. Loubeyre P, Revel D, Douek P, Delignette A, Baldy C, et al. (1994) Dynamic contrast-enhanced MR angiography of pulmonary embolism: comparison with pulmonary angiography. *AJR Am J Roentgenol* 162: 1035-1039.
97. Magazzini S, Vanni S, Toccafondi S, Paladini B, Zanobetti M, et al. (2007) Duplex ultrasound in the emergency department for the diagnostic management of clinically suspected deep vein thrombosis. *Acad Emerg Med* 14: 216-220.
98. Mathis G, Bitschnau R, Gehmacher O, Scheier M, Kopf A, et al. (1999) Chest ultrasound in diagnosis of pulmonary embolism in comparison to helical CT. *Ultraschall Med* 20: 54-59.

99. Mathis G, Blank W, Reissig A, Lechleitner P, Reuss J, et al. (2005) Thoracic ultrasound for diagnosing pulmonary embolism: a prospective multicenter study of 352 patients. *Chest* 128: 1531-1538.
100. Mathis G, Metzler J, Feurstein M, Fussenegger D, Sutterlutti G (1990) [Lung infarcts detected with ultrasonography]. *Ultraschall Med* 11: 281-283.
101. Mathis G, Metzler J, Fussenegger D, Sutterlutti G, Feurstein M, et al. (1993) Sonographic observation of pulmonary infarction and early infarctions by pulmonary embolism. *Eur Heart J* 14: 804-808.
102. Mayo JR, Remy-Jardin M, Muller NL, Remy J, Worsley DF, et al. (1997) Pulmonary embolism: prospective comparison of spiral CT with ventilation-perfusion scintigraphy. *Radiology* 205: 447-452.
103. Meaney JF, Weg JG, Chenevert TL, Stafford-Johnson D, Hamilton BH, et al. (1997) Diagnosis of pulmonary embolism with magnetic resonance angiography. *N Engl J Med* 336: 1422-1427.
104. Merminod T, Pellicciotta S, Bounameaux H (2006) Limited usefulness of D-dimer in suspected deep vein thrombosis of the upper extremities. *Blood Coagul Fibrinolysis* 17: 225-226.
105. Miniati M, Monti S, Pratali L, Di Ricco G, Marini C, et al. (2001) Value of transthoracic echocardiography in the diagnosis of pulmonary embolism: results of a prospective study in unselected patients. *Am J Med* 110: 528-535.
106. Miniati M, Pistolesi M, Marini C, Di Ricco G, Formichi B, et al. (1996) Value of perfusion lung scan in the diagnosis of pulmonary embolism: results of the Prospective Investigative Study of Acute Pulmonary Embolism Diagnosis (PISA-PED). *Am J Respir Crit Care Med* 154: 1387-1393.
107. Miron MJ, Perrier A, Bounameaux H, de Moerloose P, Slosman DO, et al. (1999) Contribution of noninvasive evaluation to the diagnosis of pulmonary embolism in hospitalized patients. *Eur Respir J* 13: 1365-1370.
108. Mohn K, Quiot JJ, Nonent M, Lacut K, Le Gall G, et al. (2003) Transthoracic sonography of the lung and pleura in view of a suspected pulmonary embolism: a pilot study. *J Ultrasound Med* 22: 673-678; quiz 680-671.
109. Mukherjee D, Andersen CA, Sado AS, Bertoglio MC (1991) Use of light reflection rheography for diagnosis of axillary or subclavian venous thrombosis. *Am J Surg* 161: 651-656.
110. Nellen V, Sulzer I, Barizzi G, Lammle B, Alberio L (2012) Rapid exclusion or confirmation of heparin-induced thrombocytopenia: a single-center experience with 1,291 patients. *Haematologica* 97: 89-97.
111. Musset D, Parent F, Meyer G, Maitre S, Girard P, et al. (2002) Diagnostic strategy for patients with suspected pulmonary embolism: a prospective multicentre outcome study. *Lancet* 360: 1914-1920.
112. Nilsson T, Soderberg M, Lundqvist G, Cederlund K, Larsen F, et al. (2002) A comparison of spiral computed tomography and latex agglutination D-dimer assay in acute pulmonary embolism using pulmonary arteriography as gold standard. *Scand Cardiovasc J* 36: 373-377.
113. Ost D, Rozenshtein A, Saffran L, Snider A (2001) The negative predictive value of spiral computed tomography for the diagnosis of pulmonary embolism in patients with nondiagnostic ventilation-perfusion scans. *Am J Med* 110: 16-21.
114. Oudkerk M, van Beek EJ, Wielopolski P, van Ooijen PM, Brouwers-Kuyper EM, et al. (2002) Comparison of contrast-enhanced magnetic resonance angiography and conventional pulmonary angiography for the diagnosis of pulmonary embolism: a prospective study. *Lancet* 359: 1643-1647.
115. Pascu AM, Radoi M, Coculescu M (2005) Plasma brain natriuretic peptide (BNP) increase is associated with acute right ventricular dysfunction in pulmonary embolism. *Acta Endocrinology* 1: 393-410.
116. Patel MC, Berman LH, Moss HA, McPherson SJ (1999) Subclavian and internal jugular veins at Doppler US: abnormal cardiac pulsatility and respiratory phasicity as a predictor of complete central occlusion. *Radiology* 211: 579-583.
117. Patwardhan NA, Anderson FA, Jr., Cutler BS, Wheeler HB (1983) Noninvasive detection of axillary and subclavian venous thrombosis by impedance plethysmography. *J Cardiovasc Surg (Torino)* 24: 250-255.
118. Perrier A, Bounameaux H, Morabia A, de Moerloose P, Slosman D, et al. (1996) Diagnosis of pulmonary embolism by a decision analysis-based strategy including clinical probability, D-dimer levels, and ultrasonography: a management study. *Arch Intern Med* 156: 531-536.

119. Perrier A, Desmarais S, Goehring C, de Moerloose P, Morabia A, et al. (1997) D-dimer testing for suspected pulmonary embolism in outpatients. *Am J Respir Crit Care Med* 156: 492-496.
120. Perrier A, Desmarais S, Miron MJ, de Moerloose P, Lepage R, et al. (1999) Non-invasive diagnosis of venous thromboembolism in outpatients. *Lancet* 353: 190-195.
121. Perrier A, Miron MJ, Desmarais S, de Moerloose P, Slosman D, et al. (2000) Using clinical evaluation and lung scan to rule out suspected pulmonary embolism: Is it a valid option in patients with normal results of lower-limb venous compression ultrasonography? *Arch Intern Med* 160: 512-516.
122. Perrier A, Roy PM, Aujesky D, Chagnon I, Howarth N, et al. (2004) Diagnosing pulmonary embolism in outpatients with clinical assessment, D-dimer measurement, venous ultrasound, and helical computed tomography: a multicenter management study. *Am J Med* 116: 291-299.
123. Pfeil A, Reissig A, Heyne JP, Wolf G, Kaiser WA, et al. (2010) Transthoracic sonography in comparison to multislice computed tomography in detection of peripheral pulmonary embolism. *Lung* 188: 43-50.
124. Pieralli F, Olivotto I, Vanni S, Conti A, Camaiti A, et al. (2006) Usefulness of bedside testing for brain natriuretic peptide to identify right ventricular dysfunction and outcome in normotensive patients with acute pulmonary embolism. *Am J Cardiol* 97: 1386-1390.
125. Investigators P (1990) Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 263: 2753-2759.
126. Pouplard C, Gueret P, Fouassier M, Ternisien C, Trossaert M, et al. (2007) Prospective evaluation of the '4Ts' score and particle gel immunoassay specific to heparin/PF4 for the diagnosis of heparin-induced thrombocytopenia. *J Thromb Haemost* 5: 1373-1379.
127. Prandoni P, Polistena P, Bernardi E, Cogo A, Casara D, et al. (1997) Upper-extremity deep vein thrombosis. Risk factors, diagnosis, and complications. *Arch Intern Med* 157: 57-62.
128. Pruszczyk P, Kostrubiec M, Bochowicz A, Styczynski G, Szulc M, et al. (2003) N-terminal pro-brain natriuretic peptide in patients with acute pulmonary embolism. *Eur Respir J* 22: 649-653.
129. Qanadli SD, Hajjam ME, Mesurrolle B, Barre O, Bruckert F, et al. (2000) Pulmonary embolism detection: prospective evaluation of dual-section helical CT versus selective pulmonary arteriography in 157 patients. *Radiology* 217: 447-455.
130. Quinn DA, Fogel RB, Smith CD, Laposata M, Taylor Thompson B, et al. (1999) D-dimers in the diagnosis of pulmonary embolism. *Am J Respir Crit Care Med* 159: 1445-1449.
131. Quinn RJ, Nour R, Butler SP, Glenn DW, Travers PL, et al. (1994) Pulmonary embolism in patients with intermediate probability lung scans: diagnosis with Doppler venous US and D-dimer measurement. *Radiology* 190: 509-511.
132. Reissig A, Heyne JP, Kroegel C (2001) Sonography of lung and pleura in pulmonary embolism: sonomorphologic characterization and comparison with spiral CT scanning. *Chest* 120: 1977-1983.
133. Remy-Jardin M, Remy J, Deschildre F, Artaud D, Beregi JP, et al. (1996) Diagnosis of pulmonary embolism with spiral CT: comparison with pulmonary angiography and scintigraphy. *Radiology* 200: 699-706.
134. Remy-Jardin M, Remy J, Watinne L, Giraud F (1992) Central pulmonary thromboembolism: diagnosis with spiral volumetric CT with the single-breath-hold technique--comparison with pulmonary angiography. *Radiology* 185: 381-387.
135. Rice C, Faragher J, Byyny R (2009) Duplex Ultrasound in the Emergency Department for the Diagnostic Management of Clinically Suspected Deep Vein Thrombosis. *Acad Emerg Med (Supplement SAEM Annual Meeting Abstracts)* 16: 139.
136. Rodger MA, Jones G, Rasuli P, Raymond F, Djunaedi H, et al. (2001) Steady-state end-tidal alveolar dead space fraction and D-dimer: bedside tests to exclude pulmonary embolism. *Chest* 120: 115-119.
137. Ruiz Y, Caballero P, Caniego JL, Frieria A, Olivera MJ, et al. (2003) Prospective comparison of helical CT with angiography in pulmonary embolism: global and selective vascular territory analysis. Interobserver agreement. *Eur Radiol* 13: 823-829.

138. Rumpf TH, Krizmaric M, Grmec S (2009) Capnometry in suspected pulmonary embolism with positive D-dimer in the field. *Crit Care* 13: R196.
139. Sanchez O, Wermert D, Faisy C, Revel MP, Diehl JL, et al. (2006) Clinical probability and alveolar dead space measurement for suspected pulmonary embolism in patients with an abnormal D-dimer test result. *J Thromb Haemost* 4: 1517-1522.
140. Schutgens RE, Ackermark P, Haas FJ, Nieuwenhuis HK, Peltenburg HG, et al. (2003) Combination of a normal D-dimer concentration and a non-high pretest clinical probability score is a safe strategy to exclude deep venous thrombosis. *Circulation* 107: 593-597.
141. Shields GP, Turnipseed S, Panacek EA, Melnikoff N, Gosselin R, et al. (2002) Validation of the Canadian clinical probability model for acute venous thrombosis. *Acad Emerg Med* 9: 561-566.
142. Shiver SA, Lyon M, Blaivas M, Adhikari S (2010) Prospective comparison of emergency physician-performed venous ultrasound and CT venography for deep venous thrombosis. *Am J Emerg Med* 28: 354-358.
143. Sijens PE, Oudkerk M, Berghout A, van Ingen HE, Kemperman H (2001) Comparison of a quantitative latex and a quantitative ELISA plasma D-dimer assay in the exclusion of segmental and subsegmental pulmonary embolism. *Thromb Haemost* 86: 1580-1582.
144. Sijens PE, van Ingen HE, van Beek EJ, Berghout A, Oudkerk M (2000) Rapid ELISA assay for plasma D-dimer in the diagnosis of segmental and subsegmental pulmonary embolism. A comparison with pulmonary angiography. *Thromb Haemost* 84: 156-159.
145. Sottiurai VS, Towner K, McDonnell AE, Zarins CK (1982) Diagnosis of upper extremity deep venous thrombosis using noninvasive technique. *Surgery* 91: 582-585.
146. Stanley JM, McGrath RS, Freeman MB, Stevens SL, Goldman MH (1994) Predictability of symptoms of upper extremity deep venous thrombosis in patients with central venous catheters with color duplex imaging. *Journal of Vascular Technology* 18: 71-73.
147. Stone E, Roach P, Bernard E, Briggs G, Havryk A, et al. (2003) Use of computed tomography pulmonary angiography in the diagnosis of pulmonary embolism in patients with an intermediate probability ventilation/perfusion scan. *Intern Med J* 33: 74-78.
148. Sullivan ED, Peter DJ, Cranley JJ (1984) Real-time B-mode venous ultrasound. *Journal of Vascular Surgery* 1: 465-471.
149. Sullivan ED, Reece C, Cranley JJ (1983) Phleborheography of the upper extremity. *Archives of Surgery* 118: 1134-1136.
150. Tawfik NM, Hegazy MA, Hassan EA, Ramadan YK, Nasr AS (2011) Egyptian experience of reliability of 4T's score in diagnosis of heparin induced thrombocytopenia syndrome. *Blood Coagul Fibrinolysis* 22: 701-705.
151. Teigen CL, Maus TP, Sheedy PF, 2nd, Johnson CM, Stanson AW, et al. (1993) Pulmonary embolism: diagnosis with electron-beam CT. *Radiology* 188: 839-845.
152. Theodoro D, Blaivas M, Duggal S, Snyder G, Lucas M (2004) Real-time B-mode ultrasound in the ED saves time in the diagnosis of deep vein thrombosis (DVT). *Am J Emerg Med* 22: 197-200.
153. Turkstra F, Kuijter PM, van Beek EJ, Brandjes DP, ten Cate JW, et al. (1997) Diagnostic utility of ultrasonography of leg veins in patients suspected of having pulmonary embolism. *Ann Intern Med* 126: 775-781.
154. Turnbull TL, Dymowski JJ, Zalut TE (1990) A prospective study of hand-held Doppler ultrasonography by emergency physicians in the evaluation of suspected deep-vein thrombosis. *Ann Emerg Med* 19: 691-695.
155. van Beek EJ, Kuijter PM, Buller HR, Brandjes DP, Bossuyt PM, et al. (1997) The clinical course of patients with suspected pulmonary embolism. *Arch Intern Med* 157: 2593-2598.
156. van Beek EJ, Reekers JA, Batchelor DA, Brandjes DP, Buller HR (1996) Feasibility, safety and clinical utility of angiography in patients with suspected pulmonary embolism. *Eur Radiol* 6: 415-419.

157. van Rossum AB, Treurniet FE, Kieft GJ, Smith SJ, Schepers-Bok R (1996) Role of spiral volumetric computed tomographic scanning in the assessment of patients with clinical suspicion of pulmonary embolism and an abnormal ventilation/perfusion lung scan. *Thorax* 51: 23-28.
158. van Strijen MJ, de Monye W, Schiereck J, Kieft GJ, Prins MH, et al. (2003) Single-detector helical computed tomography as the primary diagnostic test in suspected pulmonary embolism: a multicenter clinical management study of 510 patients. *Ann Intern Med* 138: 307-314.
159. Velmahos GC, Vassiliu P, Wilcox A, Hanks SE, Salim A, et al. (2001) Spiral computed tomography for the diagnosis of pulmonary embolism in critically ill surgical patients: a comparison with pulmonary angiography. *Arch Surg* 136: 505-511.
160. Verschuren F, Liistro G, Coffeng R, Thys F, Roeseler J, et al. (2004) Volumetric capnography as a screening test for pulmonary embolism in the emergency department. *Chest* 125: 841-850.
161. Verschuren F, Sanchez O, Righini M, Heinonen E, Le Gal G, et al. (2010) Volumetric or time-based capnography for excluding pulmonary embolism in outpatients? *J Thromb Haemost* 8: 60-67.
162. Vuilleumier N, Righini M, Perrier A, Rosset A, Turck N, et al. (2008) Correlation between cardiac biomarkers and right ventricular enlargement on chest CT in non massive pulmonary embolism. *Thromb Res* 121: 617-624.
163. Wallis JW, Kruij M, de Jongh-Leuvenink J, Buller HR (2000) A comparative study of two rapid D-dimer tests for the exclusion of pulmonary embolism in symptomatic patients. *Thromb Haemost* 84: 925.
164. Wells PS, Anderson DR, Rodger M, Forgie M, Kearon C, et al. (2003) Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. *N Engl J Med* 349: 1227-1235.
165. Wells PS, Anderson DR, Rodger M, Stiell I, Dreyer JF, et al. (2001) Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med* 135: 98-107.
166. Wells PS, Ginsberg JS, Anderson DR, Kearon C, Gent M, et al. (1998) Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med* 129: 997-1005.
167. Yoon YH, Lee SW, Jung DM, Moon SW, Horn JK, et al. (2010) The additional use of end-tidal alveolar dead space fraction following D-dimer test to improve diagnostic accuracy for pulmonary embolism in the emergency department. *Emerg Med J* 27: 663-667.
168. Zufferey P, Pararas C, Monti M, Depairon M (1992) Assessment of acute and old deep venous thrombosis in upper extremity by venous strain gauge plethysmography. *Vasa-Journal of Vascular Diseases* 21: 263-267.
169. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, et al. (2006) Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 6: 9.
170. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM (2013) Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* 185: E537-544.
171. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 46: 1041-1050.
172. Price CP, Christenson RH (2008) Evaluating new diagnostic technologies: perspectives in the UK and US. *Clin Chem* 54: 1421-1423.
173. Moons KG (2010) Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem* 56: 537-541.

APPENDIX

MEDLINE Search

“Heparin-induced thrombocytopenia” OR “Pulmonary embolism” OR “Venous thrombosis” OR “Lower extremity deep vein thrombosis” OR “Disseminated intravascular coagulation” AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti] OR systematic literature review [ti] OR (systematic review [tiab] AND review [pt]) OR consensus development conference [pt] OR practice guideline [pt] OR cochrane database syst rev [ta] OR acp journal club [ta] OR health technol assess [ta] OR evid rep technol assess summ [ta] OR drug class reviews [ti]) OR (clinical guideline [tw] AND management [tw])OR ((evidence based[ti] OR evidence-based medicine [mh] OR best practice* [ti] OR evidence synthesis [tiab]) AND (review [pt] OR diseases category[mh] OR behavior and behavior mechanisms [mh] OR therapeutics [mh] OR evaluation studies[pt] OR validation studies[pt] OR guideline [pt] OR pmcbook)) OR ((systematic [tw] OR systematically [tw] OR critical [tiab] OR (study selection [tw]) OR (predetermined [tw] OR inclusion [tw] AND criteri* [tw]) OR exclusion criteri* [tw] OR main outcome measures [tw] OR standard of care [tw] OR standards of care [tw]) AND (survey [tiab] OR surveys [tiab] OR overview* [tw] OR review [tiab] OR reviews [tiab] OR search* [tw] OR handsearch [tw] OR analysis [tiab] OR critique [tiab] OR appraisal [tw] OR (reduction [tw]AND (risk [mh] OR risk [tw]) AND (death OR recurrence))) AND (literature [tiab] OR articles [tiab] OR publications [tiab] OR publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR published [tiab] OR unpublished [tw] OR citation [tw] OR citations [tw] OR database [tiab] OR internet [tiab] OR textbooks [tiab] OR references [tw] OR scales [tw] OR papers [tw] OR datasets [tw] OR trials [tiab] OR meta-analy* [tw] OR (clinical [tiab] AND studies [tiab]) OR treatment outcome [mh] OR treatment outcome [tw] OR pmcbook)) NOT (letter [pt] OR newspaper article [pt] OR comment [pt]) AND ((sensitivity*[Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnose[Title/Abstract] OR diagnosed[Title/Abstract] OR diagnoses[Title/Abstract] OR diagnosing[Title/Abstract] OR diagnosis[Title/Abstract] OR diagnostic[Title/Abstract] OR diagnosis[MeSH:noexp] OR diagnostic * [MeSH:noexp] OR diagnosis, differential[MeSH:noexp] OR diagnosis[Subheading:noexp]))

EMBASE Search

heparin-induced thrombocytopenia'/exp OR pulmonary embolism'/exp OR 'venous thrombosis'/exp OR 'lower extremity deep vein thrombosis'/exp OR 'disseminated intravascular coagulation'/exp AND ([meta analysis]/lim OR [systematic review]/lim) AND

[humans]/lim AND [embase]/lim AND sensitivity AND specificity OR ('heparin-induced thrombocytopenia'/exp OR 'pulmonary embolism'/exp OR 'venous thrombosis'/exp OR 'lower extremity deep vein thrombosis'/exp OR 'disseminated intravascular coagulation'/exp AND ([meta analysis]/lim OR [systematic review]/lim) AND [humans]/lim AND [embase]/lim AND diagnostic AND accuracy)

CHAPTER 9

General Discussion

Hereditary and acquired disorders of haemostasis contribute significantly to thromboembolism and bleeding, which represents a significant health burden. Thus, improving diagnosis and treatment of haemostatic disorders will not only enhance care in a large number of patients but also help to stabilize budgets of health insurances and health care systems. Our research focuses on laboratory tests as well as clinical assessment scores, which are essential for diagnosis, prognosis and treatment of thromboembolic and bleeding disorders in numerous situations. To gain value in clinical practice, medical tests have to demonstrate validity, accuracy and precision.

Validity of diagnostic tests used in common practice

The extend of validity and accuracy problems we identified in tests that are already in common use, is remarkable. For example, thromboelastometry analysis is widely used, not only in patients with massive bleeding. However, our investigation indicated important limitations in reproducibility [1] and major factors of impact, that are frequently prevalent in patients where the analysis is typically applied [2]. Moreover, technical validity is not fully established [3]. Additionally, we found the reproducibility of the 4T's score, a commonly used clinical scoring system to rule-out heparin-induced thrombocytopenia, to be inacceptable low in real-life clinical practice [4]. Moreover, coagulation tests are often not comparable between different laboratories [5]. The mentioned issues can seriously limit the clinical value and no one can foresee possible consequences. It cannot be ruled out that decisions made on the basis of miss-interpretations, false alarms or false-negative findings can increase patients' burden.

Extend and quality of test evaluation

Our investigations also indicate that evaluations of medical and laboratory tests are often fragmentary and inadequate. It is remarkable that all tests investigated are already implemented and recommended by current guidelines. Important validity aspects of thromboelastometry analysis like reproducibility, factors of impact and comparison with the gold standard have not been adequately addressed [1-3]. Interobserver reliability of the 4T's score in real-life practice was not tested [4]. In particular, methodological quality of studies investigating a large spectrum of diagnostic tests for venous thromboembolic diseases is limited and associated with a high risk of bias due to design-related shortcomings [6]. Thus, decisions regarding diagnosis, prognosis and treatment that are based on these tests are subject to considerable uncertainty. It may also lead to wrong diagnosis, unnecessary delays in starting appropriate treatment, or even hazardous therapy.

Risks of applying medical tests

Application of a medical test is acknowledged to be an intervention that may bring significant harm to patients. The spectrum of risks is large. **Table 1** summarises them and gives examples. Misclassifications due to false-positive or false-negative test results are part of daily clinical practice. Wrong diagnoses eventually lead to application of unnecessary and risky treatments. In addition, misclassifications may result in uncertainty and initiation of further diagnostic tests that might further increase uncertainty and adverse effects without improving diagnostic value. In addition, if physicians reject a correct diagnosis on the base of false negative test results, there might be a delay or even non-observance in the diagnostic work-up. The consequences appear apparent in life-threatening situations such as pulmonary embolism. Testing has not only direct effects on health outcomes but it influences emotional well-being and social functioning, it has cognitive as well as behavioural effects. These factors may intensify the positive effects on health, but they can also counteract and result in lasting social, emotional or cognitive impairment. It appears as self-evident, that all the effects mentioned above will increase health care costs for the patient, health insurances and the public. However, experience from clinical practice suggests that they are rarely taken into account while considering a diagnostic test [7].

Table 1: Potential risks of applying medical tests to patients

| Risk | Description | Example |
|---|--|---|
| Direct adverse event | Invasive diagnostic tests may have direct adverse events. | Cerebral angiography carries a risk of permanent neurological damage [20]. |
| False diagnosis | False positive results may result in wrong diagnosis and application of unnecessary and risky treatments. | Overdiagnosis of heparin-induced thrombocytopenia (HIT) is common due to a low positive predictive value of respective antibody tests. Treatment of HIT with alternative anticoagulants is associated with a high risk of major bleedings [21]. |
| Rejection of correct diagnosis | False negative results may lead to rejection of a correct diagnosis and unnecessary delays in starting appropriate treatments. | Negative levels of D-dimers may delay diagnosis of pulmonary embolism what sets patient at risk for progression [22]. |
| Initiation of additional investigations | Uncertainty in interpretation of test results (positive or negative) may result in further investigations. | Recognising elevated D-dimers in patients with pneumonia may bring physicians to order a CT scan that carries a risk of contrast-induced nephropathy [23]. |
| Withdrawal of treatments | False positive results may lead to suspicion of certain diseases, which constitutes a contraindication for treatments of other diseases. | EDTA-dependent pseudothrombocytopenia (false positive result) may mimic thrombocytopenia and prevent physicians from treatment with anticoagulants, antiaggregants and fibrinolytic therapy [24]. |
| Increased costs | False positive test results may lead to additional investigations and/or treatments with a relevant increase of costs. | Overuse of D-dimer tests in emergency departments may trigger a large number of CT scans to exclude pulmonary embolism [23]. |
| Adverse emotional effects | Receiving a test result may lastingly affect mental health, increase anxiety, stress and depression. | Testing for human papillomavirus in women with borderline or mildly dyskaryotic smear test raises anxiety and distress and diminishes health-related quality of life [25]. |
| Adverse social effects | Results of medical tests may affect relationships and social functioning. | Women tested positive for BRCA1 and BRCA2 after having diagnosed for breast cancer report on feelings of guilt and disrupted family relationships [26]. |
| Adverse cognitive effects | Receiving a test result influences patients' beliefs, perceptions and understanding of their condition what affects patients' adaptive behaviour. | Differences in patients' perceptions are associated with outcomes in patients with chronic illnesses including self-management behaviours and quality of life [27]. |
| Adverse behavioural effects | Test results affect risk perceptions and anxiety what influences patients' behaviour, for example with regard to adherence to follow-ups, investigations and treatments as well as preventive lifestyle. | Results of genetic testing for BRCA1 and BRCA2 affects adherence to follow-up investigations and treatments [28]. |

Risks of premature implementation

Implementing medical tests prematurely, without taking into account all aspects of test evaluation, may lead to an excess in costs, wrong diagnosis, unnecessary delays in starting appropriate treatment, or even hazardous therapy [8]. The dexamethasone suppression test is an example where the initial phases of evaluation haven't been done adequately [9]. It was introduced for the diagnosis of major depression, and disseminated rapidly after first optimistic results. It took several years until it was noticed, that the test is subject to a large bio-variability, making it unreliable for more general application [9]. Two other tests are examples of inadequate diagnostic accuracy studies: the carcinoembryonic antigen test as screening test for colon cancer and the 125I-fibrinogen leg scan as a non-invasive diagnostic test for the diagnosis of deep vein thrombosis [10,11]. Both were disseminated rapidly soon after first studies. However, conduction of adequate accuracy studies years after introduction revealed serious limitations in accuracy, yielding no diagnostic value. The value of outcome studies is displayed with the example of rapid resonance imaging (rapid MRI). Many clinicians have enthusiastically received it. However, years later a randomised controlled trial revealed no benefit with regard to clinical endpoints but an increase in costs and the number of spine operations [12]. Thus, premature implementation of tests into clinical practice may not only cause harm to our patients but can lead to significant costs to the public.

A phased approach to test evaluation

Our results call for sustained efforts to ensure a comprehensive evaluation of medical tests before implementation. Some guidelines exist with regard to diagnostic accuracy studies (STARD, QUADAS/QUADAS-2; [13,14]) but compliance is limited [6]. Not only researchers or manufacturers of analytical devices but also regulatory authorities and editors publishing these investigations should pay more attention to them. Furthermore, little attention is paid to other aspects of test evaluation. In contrast to drug development there is no structured sequence of studies required for registration or implementation [15]. There is a consensus that comprehensive studies of "analytical" or "technical" performance should precede investigations of diagnostic accuracy. However, guidance on these technical issues is limited and no formal guideline exists. Some authors suggest a phased evaluation according to FDA requirements for approval of new drugs [8,16-19]. The characteristics to establish within the individual phases are displayed in **Table 2**. A phased approach has a number of advantages: (1) all test characteristics are assigned to a particular study phase, (2) study aims are clearly defined in every phase, (3) study design of a particular phase takes the results of the previous one into account, (4) costly evaluation can be stopped after every phase if results are not appropriate, (5) it simplifies review of the literature for assessing the value of a diagnostic test. Furthermore, implementing a phased approach may stimulate registration au-

thorities, journal editors, and even responsibilities of health care providers to critically assess the value of laboratory and medical tests. Thus, implementing a phased approach for evaluating medical tests has not only the potential to improve patient care but even save relevant amount of costs to the public.

What are the implications regarding the tests mentioned in this thesis? There is a great need for clinical studies of thromboelastometry analysis taking our results regarding reproducibility and factors of impact into account. Furthermore, studies must be designed that fulfil the requirements of current guidelines. Physicians should interpret the results of thromboelastometry carefully until these studies are available and take other laboratory and clinical information into account for decision-making. Furthermore, authors of current guidelines are advised to mention the uncertainties to avoid misinterpretations in clinical practice. With regard to the 4T's score should physicians not solely rely on clinical information to rule-out heparin-induced thrombocytopenia, particularly in special settings such as critically ill patients. We already planned a study investigating the value of a diagnostic algorithm combining 4T's score and rapid available immunoassay to determine the probability of heparin-induced thrombocytopenia at bedside.

With our research, we expect to make a thoughtful contribution to the evaluation of medical tests in the area of thrombosis and haemostasis. We hope that accuracy and validity of test can be improved and decisions on patient's diagnosis, prognosis and treatment can be placed on a more secure foundation.

Table 2: Proposal for a phased evaluation: criteria to establish during evaluation of a laboratory test

| Phase of evaluation | Criteria | Explanation |
|----------------------------|--|---|
| Analytical characteristics | Pre-analytical factors | Factors that influence measurements “before” actual analysis like <i>technique of blood drawing, stability over time</i> and <i>influences of transport and sample medium</i> |
| | Factors of impact | Influences of drugs, disease conditions and other matrix effects (<i>Selectivity</i>) |
| | Biological factors | Any kind of biological variation including <i>circadian variation, age, sex, weight and pregnancy</i> |
| | Technical accuracy | Characteristics like <i>technical accuracy</i> (degree of closeness to the true value), <i>precision</i> (agreement within and between series), <i>reproducibility</i> (agreement between observers, devices, and laboratories), <i>linearity, limit of detection</i> . |
| Diagnostic accuracy | Sensitivity | True positive rate – the probability of a positive test result in all patients with the target disease |
| | Specificity | True negative rate – the probability of a negative test result in all patients without the target disease |
| | Positive predictive value | The probability of having the disease in all patients tested positive |
| | Negative predictive value | The probability of having not the disease in all patients tested negative |
| | Positive likelihood ratio | Quantifies, if a positive test results in targeted patients meaningful changes the probability of having a disease |
| | Negative likelihood ratio | Quantifies, if a negative test results in targeted patients meaningful changes the probability of having not a disease |
| Health outcomes | Relevant outcome parameters of morbidity and mortality | Outcome studies (mostly designed as randomised controlled trials) address the question whether applying the test changes mortality or relevant morbidity parameters (eg. length of stay, infection rate). |
| Cost effectiveness | Cost effectiveness | Cost effectiveness analyses investigate whether applying the test is cost-effectiveness for the health care system compared with alternative pathways. |
| Decisions | Decision making algorithms | Decision-making analyses develop algorithms and guidelines for decisions when and how to use the medical test within diagnostic and treatment pathways. |
| Organisational impact | Changes in health care | Has implementation of the test changed health care in a specific setting? |

Adapted from [8,16-18].

References

1. Nagler M, ten Cate H, Kathriner S, Casutt M, Bachmann LM, et al. (2014) Consistency of thromboelastometry analysis under scrutiny: Results of a systematic evaluation within and between analysers. *Thromb Haemost* 30: 1161-1166.
2. Nagler M, Kathriner S, Bachmann LM, Wuillemin WA (2013) Impact of changes in haematocrit level and platelet count on thromboelastometry parameters. *Thromb Res* 131: 249-253.
3. Nagler M, Ivic S, Nil S, ten Cate H, Bachmann LM, et al. (2013) Technical efficacy of thromboelastometry and thrombelastography analysis in the evaluation of haemostasis: A Systematic Review. (submitted for publication).
4. Nagler M, Fabbro T, Wuillemin WA (2012) Prospective Evaluation of the Inter-Observer Reliability of the 4T's Score in Patients with Suspected HIT. *Journal of thrombosis and haemostasis: JTH* 10: 151-152.
5. Nagler M, Bachmann LM, Alberio L, Angelillo-Scherrer A, Asmis LM, et al. (2013) Variability between laboratories performing coagulation tests with identical platforms: a nationwide evaluation study. *Thromb J* 11: 6.
6. Nagler M, Bossuyt PM, ten Cate H, Wuillemin WA, Faes L, et al. (2014) Methodological shortcomings and design-related bias in studies investigating diagnostic tests for venous thromboembolic diseases. (submitted for publication).
7. Hood VL, Weinberger SE (2012) High value, cost-conscious care: an international imperative. *Eur J Intern Med* 23: 495-498.
8. Lijmer JG, Leeflang M, Bossuyt PM (2009) Proposals for a phased evaluation of medical tests. *Med Decis Making* 29: E13-21.
9. Nierenberg AA, Feinstein AR (1988) How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 259: 1699-1702.
10. Fletcher RH (1986) Carcinoembryonic antigen. *Ann Intern Med* 104: 66-73.
11. Lensing AW, Hirsh J (1993) 125I-fibrinogen leg scanning: reassessment of its role for the diagnosis of venous thrombosis in post-operative patients. *Thromb Haemost* 69: 2-7.
12. Jarvik JG, Hollingworth W, Martin B, Emerson SS, Gray DT, et al. (2003) Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 289: 2810-2818.
13. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155: 529-536.
14. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy. Clin Chem* 49: 1-6.
15. Price CP, Christenson RH (2008) Evaluating new diagnostic technologies: perspectives in the UK and US. *Clin Chem* 54: 1421-1423.
16. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 46: 1041-1050.
17. Moons KG (2010) Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem* 56: 537-541.
18. Bossuyt PM, Reitsma JB, Linnet K, Moons KG (2012) Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 58: 1636-1643.
19. Gazelle GS, Kessler L, Lee DW, McGinn T, Menzin J, et al. (2011) A framework for assessing the value of diagnostic imaging in the era of comparative effectiveness research. *Radiology* 261: 692-698.
20. Cloft HJ, Joseph GJ, Dion JE (1999) Risk of cerebral angiography in patients with subarachnoid hemorrhage, cerebral aneurysm, and arteriovenous malformation: a meta-analysis. *Stroke* 30: 317-320.
21. Cuker A (2011) Heparin-induced thrombocytopenia (HIT) in 2011: An epidemic of overdiagnosis. *Thrombosis and haemostasis* 106: 993-994.

22. Aydogdu M, Dogan NO, Sinanoglu NT, Oguzulgen IK, Demircan A, et al. (2013) Delay in diagnosis of pulmonary thromboembolism in emergency department: is it still a problem? *Clin Appl Thromb Hemost* 19: 402-409.
23. Crichlow A, Cuker A, Mills AM (2012) Overuse of computed tomography pulmonary angiography in the evaluation of patients with suspected pulmonary embolism in the emergency department. *Acad Emerg Med* 19: 1219-1226.
24. Kocum TH, Katircibasi TM, Sezgin AT, Atalay H (2008) An unusual cause of mismanagement in an acute myocardial infarction case: pseudothrombocytopenia. *Am J Emerg Med* 26: 740 e741-742.
25. Maissi E, Marteau TM, Hankins M, Moss S, Legood R, et al. (2005) The psychological impact of human papillomavirus testing in women with borderline or mildly dyskaryotic cervical smear test results: 6-month follow-up. *Br J Cancer* 92: 990-994.
26. Metcalfe KA, Narod SA (2002) Breast cancer risk perception among women who have undergone prophylactic bilateral mastectomy. *J Natl Cancer Inst* 94: 1564-1569.
27. Petrie KJ, Jago LA, Devcich DA (2007) The role of illness perceptions in patients with medical conditions. *Curr Opin Psychiatry* 20: 163-167.
28. Lerman C, Hughes C, Croyle RT, Main D, Durham C, et al. (2000) Prophylactic surgery decisions and surveillance practices one year following BRCA1/2 testing. *Prev Med* 31: 75-80.

Summary

Disorders of haemostasis contribute significantly to thromboembolism and bleeding, which is a relevant part of the world's healthcare burden. Laboratory tests as well as clinical assessment scores are essential for diagnosis, prognosis and treatment of thromboembolic and bleeding disorders. However, it is acknowledged that the use of any medical test is an intervention that may bring significant harm to patients. For being of value in clinical practice, tests have to demonstrate validity, accuracy and reliability. A number of authors revealed serious limitations in knowledge of validity, accuracy and reliability as well as clinical value of several tests used in the work-up of patients with a bleeding tendency or thromboembolic disorders.

The research presented in this thesis should be seen as a contribution to this problem. The thesis comprises a series of projects, which were performed at Luzerner Kantonsspital, Lucerne, Switzerland in cooperation with several tertiary haematological laboratories in Switzerland and the Cardiovascular Research Institute, Maastricht University Medical Center, The Netherlands. We aimed to investigate validity and accuracy of laboratory parameters as well as clinical assessment tools used in a broad range of situations for evaluation of haemostasis.

In **Chapter 2**, we report about a systematic review investigating the technical efficacy of thromboelastometry (ROTEM) and thromboelastography (TEG) analysis compared with established parameters of haemostasis. Using comprehensive retrieval methods, no study of ROTEM and only four studies of TEG could be identified. Only one study reported correlations properly and revealed only a weak to moderate correlation with established parameters of haemostasis. It was shown that evidence on the technical efficacy is sparse. The relationship between measurements of ROTEM and TEG and established parameters of haemostasis remains ill understood and appears to be extremely limited. In view of the scarcity of the data further studies providing compelling results are required to justify more upstream evaluations in specific clinical situations.

In **Chapter 3**, we present the results of a cross-sectional study investigating the consistency of thromboelastometry analysis in different situations. Large differences in the results of some parameters and lack of homogeneity were observed. Differences appeared not only between analysers, but also between the different channels of the same analyser, between morning and afternoon measurements and when measuring four weeks apart. Furthermore, there was an inconsistency within individual tests. For

the first time, discrimination could be made between parameters and situations, where thromboelastometry analysis is reproducible and others, where it is not. Thus, the results provide essential prerequisites for studies on validity of thromboelastometry analysis.

In **Chapter 4** we report on an in-vitro study investigating factors of influence on thromboelastometry analysis. We identified haematocrit level (Hct) and platelet count (PLT) as major, independent factors of influence on a broad range of parameters. Our results emphasise, that Hct and PLT may act as relevant confounding variables. The investigation suggests that readings of thromboelastometry parameters need to be adjusted by Hct and PLT in clinical practice and future investigations. Within the appendix of Chapter 4, we discuss the significance of platelet count as a confounding variable in a study employing thromboelastometry measurements in patients with myeloproliferative disorders.

Chapter 5 reports on a prospective cohort study investigating the accuracy of point-of-care (POCT) coagulometer if used by patients determining their INR level themselves. Data on 543 patients trained for patient self-management between 2006 and 2009 within the “Coagulationcare” initiative in Switzerland were available for analysis. All applied measures revealed a good level of agreement between capillary measurements by the patients and venous measurements by certified laboratories using an established reference method. These results confirmed that POCT coagulometers are reliable instruments and of value in patient self-management of oral anticoagulation.

In **Chapter 6** we investigated the interobserver reliability of a clinical assessment score in real-life practice. Three different physicians rated the 4Ts score, the most common clinical assessment score for patients with suspected heparin-induced thrombocytopenia (HIT) in consecutive patients evaluated for suspected HIT. The investigation showed not only a limited agreement of different observers but systematic differences between different observer groups. This may limit application of results of validation studies of 4Ts score to clinical practice and warrants further specification of items of 4Ts score. The appendix of Chapter 6 discusses the value of the 4T’s score in critically ill patients in view of a large diagnostic accuracy study

In **Chapter 7** we present the results of a multicentre evaluation study investigating the variability of laboratory tests between different laboratories. Data from eight laboratories measuring fibrinogen twice in twenty healthy subjects with one out of 3 different platforms and single measurements of PT, and coagulation factors II, V, VII, VIII, IX, X, XI and XIII were analysed. Variance components that could be attributed to technicians or laboratory procedures were substantial, led to disappointingly low intraclass correlation coefficients for several factors and were pronounced for some of the platforms. The results indicate that variability of coagulation parameters between specialised laboratories is substantial despite state-of-the-art coagulometers and reagents,

great efforts to guarantee precision of laboratory tests and external quality assessment programmes. The findings call for sustained effort to raise the level of standardisation of structures and procedures involved in the quantification of coagulation factors.

Chapter 8 demonstrates the results of a literature survey. We assessed the methodological quality of studies evaluating diagnostic tests for venous thromboembolic diseases and quantified the direction and impact of biased studies on average diagnostic performance. Diagnostic systematic reviews in five target conditions were searched; two-by-two tables and the results of the reviews' quality assessment for each primary study were extracted. A meta-analysis, comparing the pooled diagnostic odds ratios of studies unlikely to be biased with studies likely to be biased according to the QUADAS-2 quality items was performed. It was found that studies investigating diagnostic accuracy of tests used to identify venous thromboembolic diseases are flawed by methodological shortcomings. They are unlikely to report precise point estimates of test performance and are often likely to be biased according to QUADAS-2. Biased studies tended to exaggerate test performance. Recommended guidelines to design and report diagnostic tests should be systematically implemented.

In conclusion, disorders of haemostasis lead to thromboembolism and bleeding, two eminent clinical challenges with substantial consequences for healthcare and societal burden. Decisions on diagnosis, prognosis and treatment depend on laboratory and other medical tests, which must be valid, accurate and precise to be of clinical value. Premature application in clinical practice may lead to wrong diagnosis, unnecessary delays in starting appropriate treatment, or even hazardous therapy. The research presented in this thesis identified major problems in precision, accuracy and validity of tests used in the work-up of patients with a bleeding tendency and thromboembolic disorders. Our results call for concerted efforts to ensure appropriate evaluation studies when evaluating medical tests. Furthermore, established guidelines to design and report diagnostic tests should be systematically adopted in the diagnostic literature. The needs we identified in this thesis may promote a large number of studies, which may bring the development of medical tests further, enhance validity, accuracy and precision and subsequently improve patient care.

Samenvatting

Stoornissen in de hemostase dragen in belangrijke mate bij aan het ontstaan van trombo-embolieën en bloedingen, hetgeen een belangrijke druk legt op de wereldwijde gezondheidszorg. Zowel laboratoriumonderzoeken als klinische beoordelingsscores zijn essentieel voor diagnose, prognose en behandeling van deze aandoeningen. Echter, het is onderkend dat gebruik van medische tests de patiënt mogelijk schade toebrengt. Om van belang te kunnen zijn in de klinische praktijk moeten tests valide, accuraat en betrouwbaar zijn. Een aantal auteurs heeft ernstige beperkingen aan het licht gebracht met betrekking tot kennis van validiteit, nauwkeurigheid en betrouwbaarheid alsmede klinische waarde van tests die gebruikt worden in de behandeling van patiënten die een bloedingsneiging of trombo-embolieën hebben.

Het hier gepresenteerde onderzoek zou gezien moeten worden als een bijdrage aan de oplossing van dit probleem. Dit proefschrift beschrijft een serie projecten, die werden uitgevoerd bij het Luzerner Kantonsspital, Lucerne, Zwitserland en het hart- en vaatcentrum, Maastricht University Medical Center+, Nederland. Voornaamste doelstelling was de validiteit en betrouwbaarheid te beoordelen van laboratorium tests en klinische beoordelingsinstrumenten die in vele situaties worden ingezet ter beoordeling van de hemostase.

In hoofdstuk 2 rapporteren we de resultaten van een systematische review, waarin de technische effectiviteit van trombo-elastometrie (ROTEM) en trombo-elastografie (TEG) analyses worden vergeleken met de 'gouden standaard'. De relatie tussen TEG en ROTEM metingen en de 'gouden standaard' zijn zeer beperkt en bovendien onduidelijk. Verdere studies die overtuigende resultaten zullen opleveren zijn nodig om tot een betere evaluatie te komen.

In hoofdstuk 3 presenteren we de resultaten van van een cross-sectionele studie waarin de trombo-elastometrie analyses in verschillende situaties wordt beoordeeld. Grote verschillen in resultaten van enkele parameters en een gebrek aan homogeniteit werden vastgesteld. Er bleken niet alleen verschillen te bestaan tussen verschillende apparaten, maar zelfs tussen verschillende kanalen binnen hetzelfde apparaat, tussen ochtend- en middagmetingen en tussen metingen uitgevoerd in periodes van 4 weken. Bovendien was er inconsistentie binnen individuele tests. Voor de eerste keer kon er een onderscheid worden gemaakt tussen parameters en situaties, waarin trombo-elastometrie analyses reproduceerbaar zijn en waarin niet. De resultaten leveren dus

belangrijke voorwaarden voor studies die de validiteit van trombo-elastometrie als onderwerp hebben.

In hoofdstuk 4 rapporteren we de resultaten van een in vitro studie waarin de factoren die van invloed zijn op de trombo-elastometrie metingen onderzocht worden. We ontdekten dat hematocrietwaarden (Hct) en trombocytenaantal (PLT) onafhankelijke determinanten zijn van een groot aantal parameters. Onze bevindingen suggereren dat bepalingen van trombo-elastometrie metingen moeten worden bijgesteld op grond van Hct en PLT, zowel in de klinische praktijk als in toekomstige klinische studies. In de appendix van dit hoofdstuk bespreken we het belang van PLT als confounder in een studie waarin trombo-elastometrie metingen worden uitgevoerd in patiënten die lijden aan myeloproliferatieve aandoeningen.

In hoofdstuk 5 wordt een prospectieve cohort studie besproken waarin de nauwkeurigheid van de point of care (POCT) coagulometer beoordeeld wordt als deze gebruikt wordt door patiënten om hun INR te bepalen. Uit alle metingen bleek dat er overeenstemming bestond tussen door patiënten uitgevoerde capillaire metingen in vergelijking met door gecertificeerde laboratoria, die een erkende referentie methode hanteerden, uitgevoerde veneuze metingen. Deze resultaten bevestigen dat POCT coagulometers betrouwbaar zijn en van waarde zijn voor patiënten die hun orale antistolling met een vitamine K antagonist zelf regelen.

In hoofdstuk 6 hebben we gekeken naar de inter-observator betrouwbaarheid van klinische beoordelings scores in de dagelijkse praktijk. Drie verschillende artsen bepaalden de 4T's score, de meest gebruikte klinische beoordelingscore voor patiënten die verdacht worden van heparine geïnduceerde trombocytopenie (HIT), in opeenvolgende patiënten die werden beoordeeld op mogelijke HIT. Dit onderzoek toonde niet alleen een beperkte inter-observator betrouwbaarheid aan, maar ook systematische verschillen tussen verschillende groepen beoordelaars. Deze bevindingen kunnen de toepassingsmogelijkheden van validatiestudies van 4T's scores in de klinische praktijk beperken en rechtvaardigen verdere specificatie van onderdelen van de 4T's score. In de appendix wordt de waarde van de 4T's score in 'critically ill' patiënten in het licht van een grote diagnostische betrouwbaarheidsstudie besproken.

In hoofdstuk 7 presenteren we de resultaten van een multicentre studie, waarin de variabiliteit van laboratorium tests uitgevoerd in verschillende laboratoria wordt geëvalueerd. Variantie die zou kunnen worden toegeschreven aan technici of laboratorium procedures is substantieel en leidde tot teleurstellend lage intra-klasse correlatie coëfficiënten voor verschillende factoren; deze effecten waren ook sterk verschillend per apparaat. De resultaten geven aan dat variantie van stollingsparameters tussen gespecialiseerde laboratoria aanzienlijk is, ondanks 'state of the art' coagulometers en reagentia, forse inspanningen teneinde precisie van laboratorium testen te garanderen en externe kwaliteits beoordelingsprogramma's. Deze bevindingen vragen om hernieuwde

inspanningen met als doel verhoging van het niveau van standaardisatie van structuren en procedures.

In hoofdstuk 8 presenteren we de resultaten van een literatuuronderzoek, waarin we de methodologische kwaliteit beoordelen van studies die diagnostische tests voor veneuze trombo-embolieën evalueren. We hebben het effect van *biased* studieresultaten op de gemiddelde diagnostische prestatie berekend. Gebleken is dat studies die de diagnostische waarde van tests beoordelen van gebrekkige methodologische kwaliteit zijn. Het is niet waarschijnlijk dat deze studies precieze puntschattingen presenteren en de studieresultaten zijn vaak vertekend volgens QUADAS-2. Aanbevolen richtlijnen voor het ontwerp van diagnostische tests en rapportage hierover zouden systematisch uitgevoerd moeten worden.

Samenvattend kunnen we stellen: stoornissen in de hemostase leidend tot het ontstaan van trombo-embolieën en/of bloedingen, brengen uitdagingen met zich mee die invloed hebben op de gezondheidszorg en leiden tot belasting van de samenleving. Beslissingen die gevolgen hebben voor diagnose, prognose en behandeling zijn vaak gebaseerd op medische tests die betrouwbaar en valide dienen te zijn om van waarde te kunnen zijn voor de klinische praktijk. Vroegtijdige toepassing in de klinische praktijk kan leiden tot verkeerde diagnoses, onnodige vertraging van de juiste behandeling of zelfs gevaar voor de patiënt. In dit proefschrift worden grote problemen gesignaleerd met betrekking tot betrouwbaarheid en validiteit van diagnostische tests. De resultaten, gerapporteerd in dit proefschrift, nodigen uit tot een gezamenlijke inspanning met als doel hoogwaardige evaluatiestudies uit te voeren. Bovendien dienen richtlijnen die ontwerp en rapportage van diagnostische tests behandelen systematisch in de diagnostische literatuur opgenomen te worden. De behoeften die we in dit proefschrift hebben vastgesteld zouden het ontstaan van nieuwe studies kunnen bevorderen. Deze studies kunnen de ontwikkeling van medische tests vooruit helpen en de betrouwbaarheid en validiteit van deze tests verbeteren, zodat uiteindelijk betere patiëntenzorg mogelijk wordt.

Valorization addendum

Relevance

Hereditary and acquired disorders of haemostasis lead to thromboembolism and bleeding. Thromboembolism and bleeding contribute relevantly to the world's healthcare burden. In addition, these clinical challenges are associated with major consequences for healthcare and public health expenditures. Improving patient care will not only lead to a reduced morbidity, mortality and better quality of life in a large number of individuals, it has the potential to save costs and stabilize budgets of health insurances and health care systems.

Laboratory tests and clinical assessment scores are essential for diagnosis, prognosis and treatment of patients with thromboembolic and bleeding disorders. However, the clinical value of medical tests critically depends on validity, accuracy and precision. These issues are under-researched by the scientific community and many validity aspects remain unclear. Often, the clinical value is uncertain, even in tests that are in common use. Applying medical test without established clinical value may lead to excess in costs, wrong diagnosis, unnecessary delays in starting appropriate treatment, or even hazardous therapy. In addition, it has been recognised that sophisticated and expensive tests that are disseminated without suitable evaluations can subsequently be found to have marginal clinical value and economic benefit.

In this thesis, we aimed to investigate the validity of diagnostic tests used in a broad range of conditions in patients with haemostatic disorders. We aimed to identify tests, settings and circumstances with an impaired validity and a major uncertainty in terms of clinical value. The limitations identified can be targeted in a large number of subsequent studies; some of them are already planned by our group. Thus, it is likely that the results of our present and future studies will generate a considerable transfer to the health care system.

Innovation

The work presented in this thesis contributes to the discussion on the evaluation of medical tests, particularly in the area of patients with bleeding and thromboembolic disorders. It generated important new evidence in two main areas.

First, we identified a large extend of validity and accuracy problems in a broad range of diagnostic tests. For example, thromboelastometry analysis is already in common use, particularly in patients with a bleeding tendency. We identified important limitations in reproducibility and major factors of impact that change thromboelastometry measurements. The reproducibility of a widely recommended clinical scoring system to rule-out heparin-induced thrombocytopenia (the 4Ts score) was found to be unacceptable low in real-life clinical practice. In addition, coagulation tests are often not comparable between different laboratories.

Second, evaluations of medical tests were found to be fragmentary and inadequate in many cases. For example, technical validity of thromboelastometry analysis is not fully established and important validity aspects like reproducibility and factors of impact are insufficiently addressed. The interobserver reliability of the 4Ts score, an important clinical scoring system was not tested in clinical practice. Notably, the methodological quality of studies investigating medical tests for venous thromboembolic diseases is limited and associated with a high risk of bias due to methodological shortcomings.

Target groups

Our findings are important for several target groups.

Physicians should be aware of the limitations of diagnostic tests that are in common use. Misclassifications can bring serious harms to their patients. They are asked to critically review their decisions based on these tests and always take all information on patient's condition into account. And even more important, they are asked to think twice about the possible consequences before ordering lab tests.

Researchers ought to consider all validity aspects and appropriate methodology while studying medical tests. In addition, reporting methodological criteria and limitations properly will promote knowledge on the value of medical tests among physicians caring for respective patients. Furthermore, rational diagnostic pathways should be developed taking this knowledge into account.

Scientific societies are asked to mention all accuracy and validity aspects before giving recommendations for or against a particular test. This attitude will promote efforts to enhance the quality of test evaluations in scientific inquiry.

Finally, health care policy makers should recognise adequate evaluation studies as prerequisite for implementation of medical tests. Furthermore, rational diagnostic pathways should be requested based on these studies and implemented to improve efficient diagnosis in clinical practice. This policy has the potential to save a relevant amount of costs.

Activities

Our findings already stimulated further research projects and a continuation of our successful collaboration between Maastricht (The Netherlands), Berne, and Lucerne (Switzerland). Several projects regarding the diagnostic and prognostic value of laboratory tests in patients with bleeding and thromboembolic were planned, funding achieved and are already under way. I started a research fellowship at Maastricht University under supervision of Prof. Hugo ten Cate (Cardiovascular) and Prof. Martin Prins/ Dr. Arina ten Cate (Epidemiology). Furthermore, a comprehensive research programme including large prospective cohort studies was developed to be implemented in Switzerland after returning to Berne. This research programme was already submitted to achieve funding.

Implementation

Getting evidence into practice is an important task to gain value in clinical practice. The research programme planned includes several projects to address this issue. We will develop rational diagnostic pathways, and investigate barriers of implementation using qualitative research methods. Furthermore, we will present our results to national and international meetings and develop guidance tools to be used in primary care.

Dankwoord

This thesis is dedicated to Helene, Sophia, Emma, and Matteo, my lovely wife and children. Thank you for all your love, patience and support. I would not be here without you.

I would like to express my deepest gratitude to my long-time supervisor, friend and mentor Walter Wuillemin. Your guidance, trust and support truly changed my life.

I wish to thank my mentor Lucas Bachmann. You guided me through all the swamps and let me see the holy grail of science.

I am indebted to Hugo ten Cate for supervising me with this thesis and for inviting me to spend one year in his group in Maastricht. Our discussions are always stimulating.

Sincere thanks to Martin Prins and Arina ten Cate who enable my fellowship in Maastricht by providing epidemiological supervision. Getting introduced to epidemiology really broadens my mind.

I am grateful to my supervisor Anne Angelillo-Scherrer, who generously supports my scientific work and even covers extra duties by herself to enable the fellowship in Maastricht.

I would like to thank all other collaborators and co-authors, which supported me. These are Lorenzo Alberio, Lars M. Asmis, Patrick M. Bossuyt, Mattias Casutt, Thomas Fabbro, Livia Faes, Sandra Ivic, Silvio Kathriner, Wolfgang Korte, Adriana Mendez, Stefanie Nil, Pascale Raddatz Müller, Guido Reber, Pirmin Schmid, Hans Stricker, Dimitrios A. Tsakiris.

Thank you Tom, for putting everything into context.

Finally, I would like to express my gratitude to all our families. Your love keeps me running.

Curriculum vitae

Michael Nagler was born on March 22, 1974 in Nordhausen, Germany. After finishing secondary school in 1993, he performed his military service for two years, mainly at Central Military Hospital, Koblenz, Germany. From 1995 to 2003 he studied Medicine at the Friedrich-Schiller University of Jena and University of Hamburg, Germany. During this time he worked as a research fellow for 6 months at the Research Unit Molecular Cell Biology, University of Jena (Prof. Böhmer). Between 2000 and 2003 he worked as an instructor in laparoscopic surgery training courses (Ethicon Endo-Surgery Inc.) and as a scientific assistant for legal procedures relating to occupational health (Attorney at law Meding, Hamburg). After graduation from medical school and moving to Switzerland in 2003 he experienced a comprehensive training in General Surgery and General Internal Medicine (Saanen District Hospital; Interlaken Regional Hospital; Cantonal Hospital Wolhusen; Private Practice, Zweisimmen; Cantonal Hospital Lucerne). He gained board certification as a Specialist in General Internal Medicine in 2009. Between 2009 and 2013 Michael received speciality training for Clinical Haematology (FMH) and Laboratory Analytics (FAMH) and gained board certification in 2013. He spent periods at Cantonal Hospital Lucerne (Prof. Wuillemin), Blood Transfusion Service Berne (Dr. Fontana) and Inselspital University Hospital, Berne (Prof. Lämmle/Prof. Angelillo-Scherrer). Between 2007 and 2011 he was working in parallel on studies in medical education and occupational psychology (Institute of Medical Education, University of Berne, Prof. Beyeler). He finished his doctorate (Dr. med.) in 2011. Since 2010 he's doing patient-oriented research with a focus on laboratory methods as well as anticoagulation therapy (Prof. Wuillemin, Lucerne and Prof. Alberio, Inselspital Berne). In collaboration with Prof. H. ten Cate (Maastricht University; Netherlands), he's actually working on his PhD-thesis entitled "Validity and diagnostic value of tests used in the diagnostic work-up of haemostatic disorders", which is aimed to defend in 2014. Since July 2014, he is research fellow at Maastricht University focussing on the application of epidemiological research in the area of thrombosis and haemostasis (Prof. H. ten Cate, Dr. A. ten Cate, Prof. M. Prins,). Actually, he is participating in the Epidemiology Masters programme at Maastricht University. He was a visiting fellow at Division of Haematology and Haemostaseology, Medical University of Vienna (Prof. Pabinger, Prof. Kyrle) in 10-11/2011. Michael is married and has three children.

Publications

Original publications (peer reviewed)

- 1 **Nagler, M.**, L.M. Bachmann, P. Schmid, P. Raddatz Müller, and W. A. Wuillemin (2014). "Patient self-management of oral anticoagulation with vitamin K antagonists in everyday practice: Efficacy and safety in a nationwide long-term prospective cohort study" *PLoS ONE* 9(4): e95761
- 2 **M. Nagler**, H. ten Cate, S. Kathriner, M. Casutt, L.M. Bachmann, W.A. Wuillemin (2014). "Consistency of thromboelastometry analysis under scrutiny: Results of a systematic evaluation within and between analysers" *Thrombosis and Haemostasis* 30;111(6): 1161–1166
- 3 **Nagler, M.**, L. M. Bachmann, L. Alberio, A. Angelillo-Scherrer, L. M. Asmis, W. Korte, A. Mendez, G. Reber, H. Stricker, D. A. Tsakiris and W. A. Wuillemin (2013). "Variability between laboratories performing coagulation tests with identical platforms – a nationwide evaluation study" *Thrombosis Journal* 11:6
- 4 **Nagler, M.**, S. Kathriner, L. M. Bachmann and W. A. Wuillemin (2013). "Impact of changes in haematocrit level and platelet count on thrombelastometry parameters" *Thrombosis Research* 131(3): 249-253
- 5 **Nagler, M.**, P. Raddatz Müller, P. Schmid, L. M. Bachmann and W. A. Wuillemin (2013). "Accuracy of the point-of-care coagulometer CoaguChek XS in the hands of patients" *Journal of thrombosis and haemostasis: JTH* 11(1): 197-199
- 6 **Nagler, M.**, S. Feller and C. Beyeler (2012). "Retrospective adjustment of self-assessed medical competencies – noteworthy in the evaluation of postgraduate practical training courses." *GMS Zeitschrift für Medizinische Ausbildung* 29(3): Doc45
- 7 **Nagler, M.**, T. Fabbro and W. A. Wuillemin (2012). "Prospective Evaluation of the Inter-Observer Reliability of the 4T's Score in Patients with Suspected Heparin-induced Thrombocytopenia." *Journal of thrombosis and haemostasis: JTH* 10(1): 151-152.

Review articles, case reports and other publications (peer reviewed)

- 8 **Nagler, M.**, A. Angelillo-Scherrer (2014). "Diagnostic value of the 4Ts score for heparin-induced thrombocytopenia in the critically ill". *Journal of Critical Care* (epub ahead of print)

- 9 **Nagler, M.**, P. Keller, D. Siegrist, and L. Alberio (2014). "A case of EDTA-dependent pseudothrombocytopenia: Simple recognition of an underdiagnosed and misleading phenomenon" *BMC Clinical Pathology* 14:19
- 10 **Nagler, M.** and H. ten Cate (2014). "Thromboelastometry changes in myeloproliferative neoplasms-surrogate for a procoagulant haemostatic imbalance or a consequence of technical reasons?" *Annals of Hematology* 2:1633–1639
- 11 Simeon, L., **M. Nagler**, and W.A. Wuillemin. "New oral anticoagulants – influence on coagulation tests" (2014) *Deutsche Medizinische Wochenschrift* 1390: 94–98
- 12 **Nagler, M.** and W. A. Wuillemin (2012). "Laboratory diagnostic with regard to new anticoagulants – monitoring and influence on coagulation tests" *Therapeutische Umschau* 69(11): 650-6.
- 13 **Nagler, M.**, M. Haslauer and W. A. Wuillemin (2011). "Fondaparinux-data on efficacy and safety in special situations." *Thrombosis research* 129(4): 407-17
- 14 **Nagler, M.**, M. Haslauer and W. A. Wuillemin (2011). "Long-term anticoagulation with fondaparinux in a patient with a mechanical heart valve." *Annals of hematology* 90(10): 1225-1226.
- 15 **Nagler, M.**, P. Erne, R. Babst, W. Korte and W. A. Wuillemin (2011). "Periinterventionselles Management der Antikoagulation und Antiaggregation." *Schweizerisches Medizin Forum* 11(23-24): 407-412.
- 16 **Nagler, M.**, M. Gregor and W. A. Wuillemin (2011). "Iron chelation with deferasirox in two patients with HFE hemochromatosis and chronic anemia." *Acta Haematologica* 126(2): 119-121.
- 17 **Nagler, M.** and W. A. Wuillemin (2011). "Leeching as a substitute for phlebotomy." *British journal of haematology* 153(4): 420.
- 18 **Nagler, M.**, B. Muller, V. Briner and R. Winterhalder (2009). "Severe hyperkalemia and bilateral adrenal metastasis." *Journal of oncology* 2009: 831979.

Other publications

- 19 Mendez, A. and **M. Nagler** (2014). "Neue orale Antikoagulantien" *PIPETTE – Swiss laboratory medicine* (ahead of print)
- 20 Simeon, L., **M. Nagler**, and W.A. Wuillemin. "Neue orale Antikoagulantien – Einfluss auf Gerinnungstests" (2014) *BDI aktuell* 3: 11-13
- 21 Wuillemin, W.A. and **M. Nagler** (2012). "Patienten-Selbstkontrolle der oralen Antikoagulation, eine POCT-Anwendung der besonderen Art" *PIPETTE – Swiss laboratory medicine* 5: 11-12

Currently under review

- * **Nagler, M.**, P. M. Bossuyt, H. ten Cate, W. A. Wuillemin, L. Faes, L. M. Bachmann. "Methodological shortcomings and design-related bias in studies investigating diagnostic tests for venous thromboembolic diseases – Literature survey"
- * **Nagler, M.**, S. Ivic, S. Nil, H. ten Cate, L.M. Bachmann, W.A. Wuillemin. "Technical efficacy of thromboelastometry and thrombelastography analysis in the evaluation of haemostasis: A Systematic Review"
- * Simeon, L., **M. Nagler**, T. Hodel, W.A. Wuillemin, and L. Alberio. "Fatal heparin-induced thrombocytopenia after "minimal" heparin exposure"
- * Colucci, G., **M. Nagler**, and L. Alberio. "Practical Use of Argatroban and Bivalirudin in Patients with Heparin-induced Thrombocytopenia"

Oral presentations

55th Annual Meeting of the "Society of Thrombosis and Haemostasis Research (GTH)", St. Gallen, Switzerland; 2012

Poster presentations

XXVI Congress of the International Society on Thrombosis and Haemostasis (ISTH), Amsterdam, The Netherlands; 2013

56th Annual Meeting of the "Society of Thrombosis and Haemostasis Research (GTH)", Munich, Germany; 2013

55th Annual Meeting of the "Society of Thrombosis and Haemostasis Research (GTH)", St. Gallen, Switzerland; 2012

54th Annual Meeting of the "Society of Thrombosis and Haemostasis Research (GTH)", Wiesbaden, Germany; 2011

Grants and Prizes

ISTH 2007 Presidential Fund educational grant (30,000 CHF); 2013

Award for the best patient-oriented project. University of Berne "Day of Clinical Research" (2,000 CHF); 2013

Best abstract price. 55th meeting of the "Society of Thrombosis and Haemostasis Research (GTH)", St. Gallen, Switzerland; 2012

Delegate of the Thrombosis Expert Academy – TEA; 2012 – 2014

References

Prof. Dr. Dr. med. W. A. Wuillemin, Head of Division of Haematology and Central Haematology Laboratory, Medical Department, Cantonal Hospital Lucerne, Lucerne, Switzerland, Phone: +41 41 205 5147, walter.wuillemin@luks.ch

Prof. Dr. med. A. Angelillo-Scherrer, Director and Head of Division of Haematology and Central Haematology Laboratory, Inselspital University Hospital, Berne, Switzerland, Phone: +41 31 632 96 01, anne.angelillo-scherrer@insel.ch