Analytical quality control of INR measurements in primary care

Anne Vegard Stavelin

Dissertation for the degree philosophiae doctor (PhD)
at the University of Bergen

2013

Department of Global Public health and Primary Care

Faculty of Medicine and Dentistry

University of Bergen

Bergen, Norway

Norwegian quality improvement of primary care laboratories

Bergen, Norway

Acknowledgements

First and foremost, I would like to thank my supervisors and mentors, ***Sverre Sandberg*** and ***Per Hyltoft Petersen***. Without you this work would not have been possible. I am honored and grateful to have worked so closely with you, and if I have achieved even a tiny part of your enormous knowledge I will be happy. You are both such an inspiration for me. I would like to thank you Sverre for your ability to always see the big picture and to put all the details into the right context. I am particularly grateful for all our interesting discussions about the meaning of work and the meaning of life. I would like to thank you Per for always coming up with exiting new ideas on how to analyze and present the data, and for bringing me into the mystery of statistics. You never stop impressing me with all your knowledge – of almost everything.

I thank all my colleagues at Noklus. I am thankful for working in such an inspiring environment with such great people. A special thanks to ***Nina Gade Christensen*** for being an inspiration for me over several years. Thank you for giving me the right amount of responsibility and challenges, and for the encouragement to take this degree. I thank ***Una Sølvik*** for good collaboration with several projects, ***Gunn Kristensen*** for giving me inspiration to climb the phd mountain, ***Grete Monsen*** for many interesting discussions and ***Ann-Helen Kristoffersen*** for inspiring cooperation with several coagulation projects and for guiding me in the map of haemostasis. Many thanks also to all ***laboratory consultants*** for good collaboration over several years and for playing an important role in selecting suitable primary care laboratories.

I would like to thank ***Anne-Margrethe Holmefjord*** and the staff at the laboratory at Haraldsplass Diaconess Hospital for participating in the studies, and for always being so positive. I also thank all participating ***patients*** and ***primary care laboratories***.

I thank ***Piet Meijer*** and ***Dianne Kitchen*** for good collaboration with one of the studies.

Last but not least, I thank my ***family and friends*** who have cheered me on and given me support and comfort when life was difficult, and for all the fun we have together. A special thank to ***my mother*** who always believes in me – in every aspects of life.

Bergen, June 2013

Anne Vegard Stavelin

"If I have seen further
it is by standing on the shoulders of giants"

Sir Isaac Newton

To my supervisors

Foreword

Decentralized laboratory testing has increased over the last decade and several different portable instruments have been developed. Such testing is termed point-of-care (POC) testing or near-patient testing. POC instruments are easy to use and are commonly used by personnel with little or no laboratory experience. The testing can be performed in many different settings, such as hospital clinics, ambulances, nursing homes, general practitioners, pharmacies, oil platforms, prisons and by patients at home for self-testing. The main objective in such testing is to produce rapid results to ensure effective clinical decision making and better patient care.

The main objective in oral anticoagulation treatment with warfarin is to prevent thromboembolic events. This treatment is, however, associated with serious side effects and correct medical dose is essential. The risk of severe bleeding is increased if the patient is overdosed and the risk of thrombosis is increased when the patient is under dosed. There is high variability in dose response among patients and frequent laboratory monitoring is therefore necessary. The laboratory test is called prothrombin time and is expressed as International Normalized Ratio (INR). It is important that the INR instruments are reliable and that the personnel perform the test correctly because the dose given depends on the test result. Ideally, the INR result should be independent of the instrument used and independent of whether the measurement is performed in a hospital or in a primary care setting. Thus, the harmonization between methods should be good, but this is, however, not always the case for INR methods.

Most of the patients on oral anticoagulation are treated in primary care. About 1700 primary care laboratories in Norway perform POC INR testing. The primary care laboratories control their instruments by performing internal quality control and external quality assessment. There are, however, many challenges in performing analytical quality control of these instruments and the control systems need to be evaluated and improved. This thesis aims to address these issues for POC INR methods, and some of the findings will also apply for POC methods in general.

List of publications

Paper I: Stavelin A, Petersen PH, Solvik U, Sandberg S. Internal quality control of prothrombin time in primary care: Comparing the use of patient split samples with lyophilized control materials. Thromb Haemost 2009;102:593-600.

Paper II: Stavelin A, Meijer P, Kitchen D, Sandberg S. External quality assessment of point-of-care international normalized ratio (INR) testing in Europe. Clin Chem Lab Med 2012;50:81-88.

Paper III: Stavelin A, Petersen PH, Solvik UO, Sandberg S. External quality assessment of point-of-care methods: Model for combined assessment of method bias and single participant performance by the use of native patient samples and non-commutable control materials. Clin Chem 2013;59:363-371.

The publications are reprinted with permission from the scientific journals.

Abstract

In Norway, most patients on oral anticoagulation with warfarin are treated in primary care. The treatment is monitored with the laboratory method prothrombin time, expressed as International normalized Ratio (INR). It is important that the INR methods have good analytical quality because the treatment (medical dose) depends on the INR result. Overdosing can cause severe bleedings and under dosing can lead to thrombosis. The laboratories in primary care control the analytical quality of their INR methods by performing internal quality control (IQC) and external quality assessment (EQA). There are, however, some challenges regarding these quality control systems. The aim of this thesis was to evaluate and suggest improvements of the analytical quality control of INR methods used in primary care.

The primary care laboratories perform IQC mainly by two different approaches; 1) a commercial lyophilized control material is analyzed on the INR method and the result is compared with some control limits, 2) a fresh patient sample is analyzed both on the INR method and on a hospital method, and the difference between the methods is compared with some control limits. The latter approach is called split sample procedure. The primary care INR method is considered "in control" if the result is within the limits and "out of control" if the result is outside the limits (error alarm). The aim of paper I was to evaluate and compare these two IQC approaches in their ability to detect systematic errors. Power functions were created by computer simulations based on empirical data from 18 primary care laboratories using the INR methods Thrombotrack, CoaguChek S, or Hemochron Jr. Signature. The control rules 1_{2S}, 1_{3S}, exponential weighted moving average, and the deviation limits of ± 10% and ± 20% were evaluated by their probability of error detection and false alarms. The results showed that the probability of detecting systematic errors was higher when lyophilized control materials were used compared to patient split samples. The probability of false alarms was, however, the same. The conclusion in paper I was that IQC of INR methods in primary care should be performed by using control materials rather than the split sample procedure. The split sample procedure with native patient samples should be restricted to method bias estimation.

International guidelines recommend that primary care laboratories should participate in an EQA scheme whenever available. The aim of paper II was to investigate if and how the European countries provide this service for point-of-care (POC) INR methods. Thirty European countries were asked, and nineteen countries reported that they do not provide EQA schemes for POC INR methods, while 12 organizations from nine countries (Austria, Czech Republic, Denmark, Finland, Hungary, Netherlands, Norway, Switzerland and United Kingdom) reported that they offer this service. All 12 organizations answered a questionnaire regarding their schemes, and the results showed that there is a vide variation in how the schemes are organized. However, the most common is to use lyophilized control materials, establish peer group target values, use an acceptability limit of 15% and distribute four samples per year. Most of the countries organize educational activities with focus on quality improvement. The study in paper II demonstrates that most European countries do not provide EQA schemes for POC INR methods, and that the disadvantages in most of the provided schemes were the use of non-commutable control materials making comparison between different POC methods impossible.

An important objective in EQA is to evaluate systematic deviations (bias) between methods. This is, however, not possible when non-commutable control materials with peer group target values are used. The aim of paper III was to develop a new EQA model in which an evaluation of method bias was incorporated in EQA schemes that use non-commutable materials. The model was developed based on the concept that a selected group of primary care laboratories should establish an estimate of the systematic deviation of the POC method from a designated comparison method by using fresh patient samples, and this information should then be incorporated in the feedback to the participants in the EQA scheme using non-commutable control materials. As a consequence, the participants will get more information about the analytical quality of their method. The model was applied twice in POC INR surveys among 1341 and 1578 participants, respectively. To estimate bias for each POC INR method, about 100 native patient samples were analyzed both by a selected group of expert primary care laboratories (72 and 69 in the first and second survey, respectively) and on a designated comparison method. Both method bias and the

deviation of a single-participant result in the EQA schemes were evaluated against separate analytical quality specifications. Two POC INR methods (CoaguChek XS Plus and Simple Simon) fulfilled the quality specification for bias, whereas one did not (Thrombotrack). More than 90% of the participants received results within the quality specification for a deviating EQA result. In conclusion, a new EQA model for POC methods was proposed in paper III. This model can be used in situations where commutable control materials are not available. An editorial in the journal *Clinical Chemistry* has recommended that EQA organizers should implement this proposed EQA model.

Abbrevations

CLSI	Clinical and laboratory standard institute
ECAA	European concerted action on anticoagulation
EQA	External quality assessment
EQALM	European organization for external quality assurance providers in laboratory medicine
EWMA	Exponentially weighted moving average
F	Coagulation factor
INR	International normalized ratio
IQC	Internal quality control
IRP	International reference preparation
ISI	International sensitivity index
ISO	International organization for standardization
JCTLM	Joint committee for traceability in laboratory medicine
MNPT	Mean normal prothrombin time
NOAC	New oral anticoagulants
Noklus	Norwegian quality improvement of primary care laboratories
P_{ED}	Probability of error detection
P_{FR}	Probability of false rejection
POC	Point-of-care
PT	Prothrombin time
TTR	Time in therapeutic range
VKA	Vitamin K antagonist
WHO	World health organization

Contents

1. INTRODUCTION

1.1 Oral anticoagulation treatment

The main objective in oral anticoagulation treatment is to prevent thrombosis. There are several types of anticoagulant drugs but the most commonly used worldwide is the vitamin K antagonist (VKA) warfarin. This drug is, however, associated with serious side effects and correct treatment is essential. Under dosing is associated with increased risk of thrombosis events and overdosing can lead to severe bleeding events. Patients with different indications for anticoagulation treatment such as atrial fibrillation or mechanical heart valves are treated on a long-term basis (often lifelong), while patients with other indications such as venous thromboembolism (e.g. deep vein thrombosis or pulmonary embolism) are treated only in short periods (usually from 6 weeks to 6 months). There is high intra- and inter-person variability in dose response due to different factors, such as age, body weight, acute and chronic diseases, diet, alcohol and drug interactions, and pharmacogenetic factors (e.g. variability in the CYP2C9 and VKORC1 gene) (1, 2). Consequently, a fixed dose is not possible and laboratory monitoring is necessary to determine the correct warfarin dose. The laboratory analysis prothrombin time, expressed as International Normalized Ratio (INR), is used to monitor the effect of the treatment. The optimal therapeutic range is 2.0 to 3.0 INR for most indications, and 2.5 to 3.5 INR for patients with mechanical heart valves (high-risk patients) (3). It has been shown that values below 1.8 INR and above 4.0 INR increase the risk of thrombosis and bleeding events, respectively (4). Time in therapeutic range (TTR) should be above 60% to achieve optimal clinical outcome (1), although others have argued that low INR variability is a more important prognostic factor than the TTR to achieve good clinical outcome (2). Consequently, it is important that the INR methods give reliable and stable values because systematic errors can lead to under or overdosing and high analytical variation can lead to unnecessary dose adjustments.

In more than 60 years, VKAs have been the only alternative of oral anticoagulants in treatment and prevention of thrombosis, but in recent years new oral anticoagulants

(NOACs), such as direct thrombin inhibitors (e.g. dabigatran) and factor Xa inhibitors (e.g. rivaroxaban and apixaban), have entered the marked. Even though these NOACs have shown promising results in clinical trials, there are many unsolved problems and VKAs will probably still play an important role in anticoagulation treatment for many years to come (1, 5).

1.2 Prothrombin time

Prothrombin time (PT) is sensitive to factor deficiency in the extrinsic and common coagulation pathway (Figure 1). The extrinsic coagulation pathway is activated when the citrated patient plasma sample is added to the reagent, which consists of thromboplastin and calcium chloride. Calcium chloride is, however, usually not present in non-citrated whole blood point-of-care (POC) methods. Thromboplastins are commercially developed products and consist of phospholipids and tissue factor extracted from different origins. Coagulation factor (F) VII in the extrinsic system and FII (prothrombin) and FX in the common pathway are dependent on vitamin K for correct synthesis. In the presence of vitamin K, the coagulation factors become γ-caboxylated. This carboxylation process and the presence of calcium ions are essential for the adherence of the clotting factors to the negatively charged phospholipids on the activated platelet surface (3). In VKA treatment, this γ-carboxylation is inhibited and coagulation factors lose their activity. Such coagulation factors are called PIVKA factors (Protein Induced in Vitamin K abscence). Thus, in VKA therapy, the blood's coagulation ability is reduced.

PT is not a measurement of a single quantitative component but a result of a coagulation cascade system. Each factor in the cascade is not quantified; instead it is the time from the activation of the cascade to the formation of stable fibrin (clot formation) that is measured.

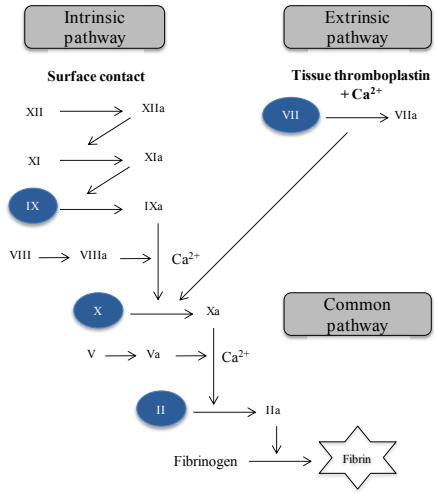

Figure 1: The coagulation cascade. The vitamin K depended coagulation factors are highlighted. Prothrombin is coagulation factor II.

1.2.1 Owren and Quick methods

In principle, there are two different methods for PT measurement. The Quick method was introduced by AJ Quick in 1935 (6) and the Owren method was introduced by PA Owren in 1947 (7). In Owren methods, FV and fibrinogen from bovine plasma are added to the reagent (so called combined reagent). Consequently, this method is sensitive only to the vitamin K-dependent factors (FII, FVII and FX) and does not reveal any deficiency in FV or fibrinogen in the patient's blood. In the Quick method, there is no added plasma (so called plain reagent), and this method is thus sensitive to FII, FVII and FX, as well as to FV and fibrinogen. This is the main difference between the Owren and Quick method but there are several variations of both methods; e.g. different degree of dilution, wet and dry chemistry, and different types of thromboplastin used. All these varieties may contribute to discrepancies in results between methods (8, 9). The Quick methods are most commonly used worldwide, whereas Owren methods are mostly used in the Nordic countries. However, it has been questioned whether Owren methods should replace Quick methods worldwide (10) because the between-laboratory variability is lower for Owren methods (9).

1.3 Standardization – the International Normalized Ratio

The standardization process of PT began already in the 1960s (11). It had been observed that different thromboplastins from different origins resulted in huge discrepancies in PT seconds, making comparisons across studies difficult. The objective was that it should be easier to compare clinical trials by introducing a common unit to report PT result. As a consequence, the INR system was introduced by the World Health Organization (WHO) in the early 1980s (12). The idea was that the PT result, expressed in INR units, should be independent of the thromboplastin by correcting for its sensitivity. Thus, the international sensitivity index (ISI) was developed in order to reduce the differences between PT methods. INR is a ratio between the patient's clotting time and the clotting time in normal plasma, corrected for the sensitivity of the thromboplastin used:

$$INR = \left(\frac{\text{sample PT (second)}}{\text{mean normal PT (second)}} \right)^{ISI}$$

The principle of the INR system is that all thromboplastins should be calibrated against an International Reference Preparation (IRP), which has an ISI equal to 1. The first IRP was established in 1976 (code 67/40) and was a combined thromboplastin reagent derived from human brain tissue (12, 13). In addition, IRPs derived from rabbit and bovine brains were established in 1978. The principle is that PT methods should be traceable to the IRP of same origin (e.g. methods using rabbit thromboplastin should be traceable to a rabbit IRP). As the IRPs were exhausted, new IRPs were established, all traceable to IRP 67/40. The current available IRPs are the recombinant human plain (coded rTF/09) and the rabbit plain (coded RBT/05) (14). No bovine IRPs are currently available, the last bovine combined (code OBT/79) is exhausted, and will not be replaced (14). It has been suggested that PT methods using bovine thromboplastins can be calibrated against a rabbit plain IRP (15).

1.3.1 Different calibration procedures

Calibration of INR means to determine the ISI and mean normal PT (MNPT) values. The original WHO calibration procedure is to analyze 60 samples from patients on VKA treatment and 20 samples from healthy individuals both with an IRP and with the thromboplastin of interest using the manual tilt-tube technique (12) (Figure 2). The PT seconds of the IRP and the thromboplastin are both log transformed and plotted on the ordinate and abscissa, respectively. An orthogonal regression equation is calculated, and the ISI value of the thromboplastin is determined as the slope multiplied with the ISI of the IRP. The MNPT value of the thromboplastin is the geometric mean of the normal plasmas.

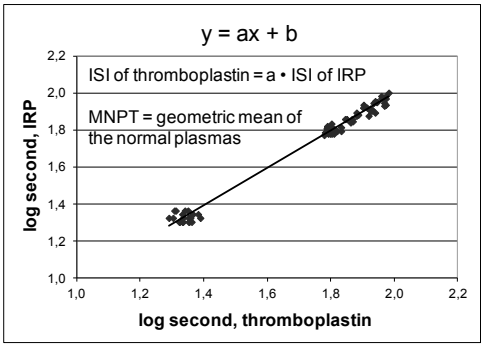

Figure 2: Example of a full WHO calibration of a thromboplastin reagent. 60 patient plasmas and 20 normal plasmas are analyzed with the thromboplastin and with an IRP by using the manual tilt-tube technique.

This calibration procedure has, however, been modified and an easier method with fewer samples are suggested (13, 16). These guidelines recommend two approaches:

1) Local ISI calibration with 20 samples from patients on VKA treatment and seven samples from healthy individuals (Figure 3). The plasmas are analyzed both with the local instrument/reagent combination and with the manual tilt-tube technique with an IRP. The ISI and MNPT values are determined as described above for the WHO procedure.

2) Direct INR determination with three abnormal and one normal lyophilized pooled plasmas (at least 10 donations in each pool) with certified INR values (Figure 4). The plasmas are analyzed with the local instrument/reagent combination and plotted against the certified values. The INR of patient samples can then be calculated or obtained directly from this calibration curve. The ISI value is equal to the slope, and the MNPT value is the anti-log of the slope divided with the negative intercept ($10^{a/-b}$).

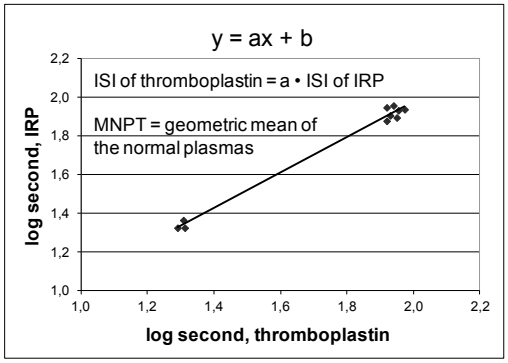

Figure 3: Example of local ISI calibration of a thromboplastin reagent. Seven patient plasmas and three normal plasmas are analyzed with the local instrument/reagent combination (x-axis) and with an IRP by using the manual tilt-tube technique (y-axis).

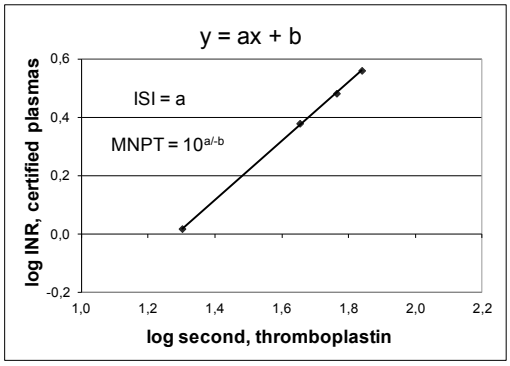

Figure 4: Example of the direct INR calibration of a thromboplastin reagent. Three abnormal pooled plasmas and one normal pooled plasma with certified INR values are analyzed with the local instrument/reagent combination.

Poller et al. (17) have compared the reliability of these two calibration procedures for different IRPs. They found that the local ISI calibration was better (i.e. gave lower between-laboratory variability) than the direct INR calibration for human and rabbit thromboplastins, whereas the direct INR calibration was the best choice for combined rabbit thromboplastins.

1.3.2 The Swedish calibrators

In the direct INR calibration, it is recommended that the pooled plasmas are certified with assigned INR values by using IRPs (16). However, a Swedish group has developed an alternative approach where fresh normal plasmas are used instead of IRPs (18). The concept is that they have developed an equation describing the relationship between PT% and INR which is traceable to an IRP, and this formula is used to create a calibration curve with diluted normal plasmas. Thus, they have established an "anchor" to the IRP hierarchy, and each batch of calibrators is assigned with certified INR values from fresh samples. According to the Swedish group, there are some clear disadvantages of using IRPs in the certification of each batch of calibrators. First, preparation of new IRPs will increase the uncertainty of the ISI value further down the hierarchy. Second, thromboplastins are more or less unstable over time, which can increase the deviation from the conventional true value. Finally, transport and storage can have influence on the stability of the IRPs. By implementing the Swedish calibrators, the between- and within laboratory variation in Swedish hospitals were reduced (19). The supplier of these calibrators is the national EQA organizer in Sweden (Equalis) and all Norwegian hospital laboratories use these calibrators, as recommended by the Norwegian INR committee (20).

The Equalis calibrators consist of only one normal and one abnormal lyophilized pooled plasma, and they are intended for three reagents only (combined rabbit thromboplastins); the SPA reagent (Diagostica Stago, France), the Nycotest PT reagent (Axis-Shield, Norway) and the Owrens PT reagent (MediRox AB, Sweden). When calibrating Quick methods, four calibrators are recommended (16) (Figure 4).

It is defendable to use only two calibrators in the direct INR determination of Owren methods because these methods are more robust than the Quick methods (18), i.e. the between-laboratory variation is lower, and the matrix effects between patient samples are lower because the degree of the dilution of the sample is higher and because FV and fibrinogen are added into the reagent.

1.3.3 Calibation hierarchy and traceability

The ISO document 17511 (21) describes the calibration hierarchy and traceability in cases where there is no traceability to SI units, no international conventional calibrator, but an international conventional reference measurement procedure (Figure 4 in the ISO document). This hierarchy is valid for some haemostatic factors (21) and it may be valid for INR methods which are calibrated by the manufacturer, e.g. POC methods (Figure 5). However, if the INR methods are calibrated by the end-user, the traceability chain is different (Figure 5). The conventional reference procedure is the manual tilt-tube technique with the use of a certain reference reagent (IRP). The calibrators are assigned with certified INR values by using fresh normal and patient samples analyzed with the reference measurement procedure (full WHO calibration). The direct INR calibration is performed by using these calibrators analysed on the local instrument/reagent combination, whereas the local ISI calibration is performed using fresh patient and normal plasmas analysed both with the instrument/reagent combination and with the manual tilt-tube technique with a specific IRP.

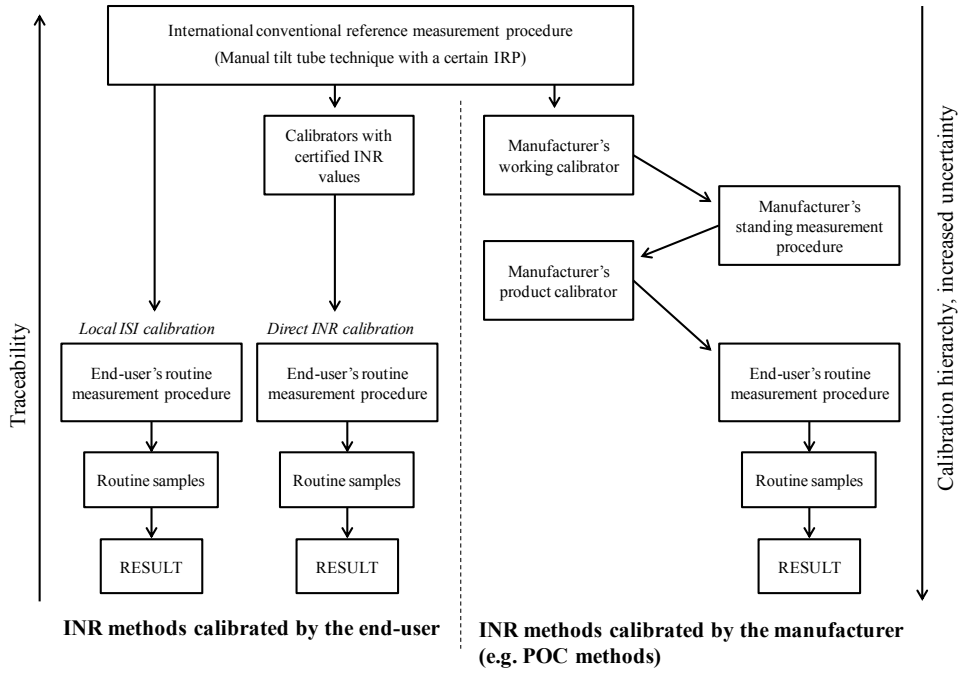

Figure 5: The calibration hierarchy and traceability of INR methods calibrated by the end-user or by the manufacturer.

1.3.4 Traceability of thromboplastins used in this thesis

As mentioned, all thromboplastins should be traceable to the first reference thromboplastin coded 67/40 (the IRP at the top of the thromboplastin hierarchy) which is defined as ISI = 1 (12). A simplified figure of the traceability chain for the different thromboplastins used in this thesis is shown in Figure 6. Users of POC INR methods cannot calibrate their own method, and the ISI and MNPT values must be provided by the manufacturer. The Hemochron Jr. Signature and CoaguChek XS Plus are traceable to the recombinant human IRP (rTF/95) (22, 23), whereas the CoaguChek S is traceable to the rabbit IRP (RBT/90) (24). The calibration values (ISI and MNPT) for each batch of reagents are incorporated in each Hemochron cuvette and in a batch specific CoaguChek code chip. The CoaguChek users must therefore change this code chip when a new batch of strips is used. The Simple Simon is

traceable to the newest rabbit IRP (RBT/05), but is in addition verified against hospital methods using the Equalis calibrators (25). This is done to optimize method harmonization between Simple Simon and the hospital methods in Sweden and Norway. All equipment needed for measurements on Simple Simon have the same batch number, and all must be replaced when a new batch is used. The calibration values are incorporated in each batch of Simple Simon instruments. The Thrombotrack instrument is used in combination with the thrombotest reagent (26). The manufacturer provides ISI and MNPT values for two different calibrations of this reagent; one traceable to the bovine IRP (OBT/79) and one traceable to the rabbit IRP (RBT/90) through the Nycotest PT reagent calibrated with the Equalis calibrators. This latter calibration is performed to optimize the harmonization of thrombotest and the hospital methods in Norway. The thrombotest users can choose which calibration curve they want to use, but in Norway it is recommended to use the "Equalis calibration" (20). The designated comparison methods (STA Compact and STA R Evolution) used in paper I and III were calibrated by the end-user using the direct INR calibration procedure with the Equalis calibrators, and are thus traceable to the rabbit IRP RBT/90.

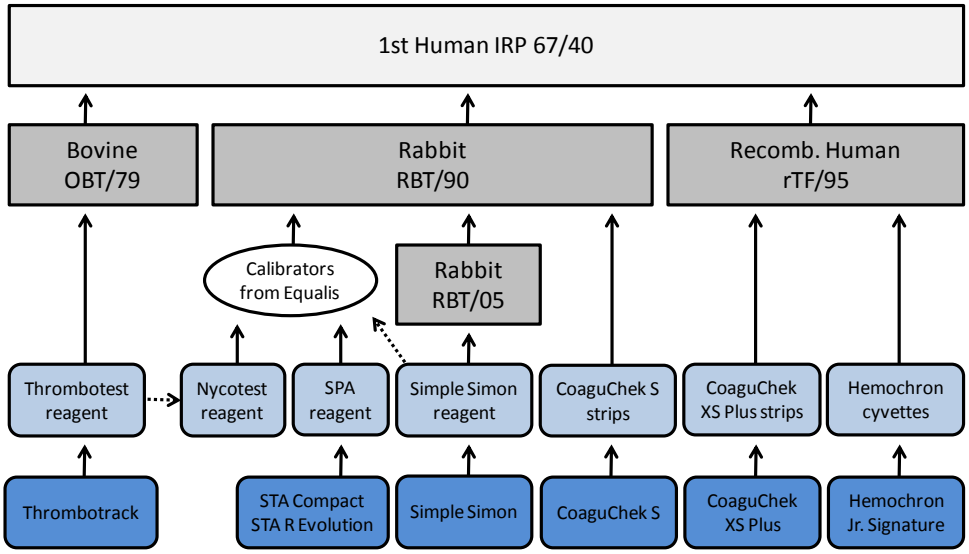

Figure 6: Simplified traceability chain of thromboplastins for the different INR instruments used in paper I and III. The reference thromboplastins (IRPs) are shown in light and dark grey, the routine tromboplastins (reagents) are shown in light blue and the instruments are shown in dark blue.

1.4 Quality in laboratory medicine

There are several factors that can influence the process from clinical findings to the diagnosing, monitoring and treatment of patients (Figure 7). This process has been called the brain-to-brain loop (27, 28) because the loop starts and ends in the clinician's brain. The first step is the pre-pre-analytical phase, in which the clinicians have to decide which laboratory tests to request based on the patient anamnesis and clinical findings. The next steps are the pre-analytical phase (e.g. patient identification, sample collection, sample stability, centrifugation, handling of sample, preparation of reagents), the analytical phase (e.g. instrumentation, analytical principle, lot number of reagent, calibration, maintenance), and the post-analytical phase (e.g. data entry, reporting and communication of results). The final step is the post-post-analytical phase, in which the clinicians should interpret the test result and decide whether and how the patient should be treated. The clinician must also

consider additional testing. The pre-analytical, analytical and post-analytical phase takes place inside the laboratories, and it is the laboratories' responsibility to assure that the quality in the different phases is good. The laboratory can be a central laboratory or a laboratory in the primary care.

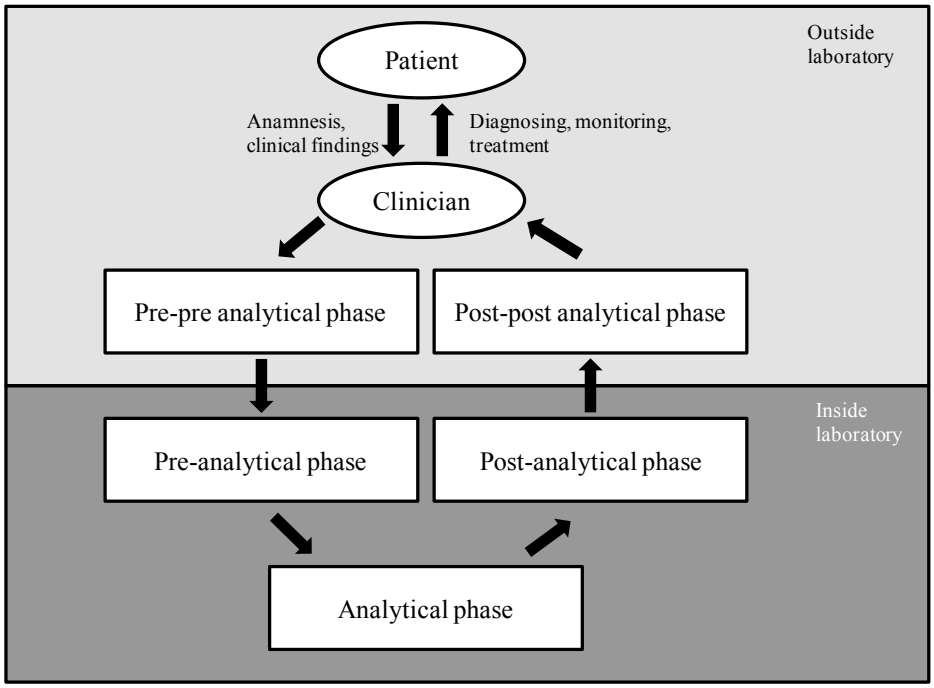

Figure 7: The brain-to-brain loop in laboratory medicine. The loop starts with the clinicians decisions of which laboratory test to request (pre-pre analytical phase) based on the clinical findings, and ends with the clinicians interpretation of the test result (post-post analytical phase) that may lead to treatment of the patient.

1.4.1 Pre-analytical phase

It has been shown that the majority of errors in laboratory medicine occur in the pre-analytical phase (29, 30). However, for POC testing it has been reported that the majority of errors occur in the analytical phase (31). Nevertheless, it is important that the pre-analytical errors are as few as possible both in POC testing and in central laboratory testing.

Pre-analytical factors in INR testing

Several pre-analytical factors may interfere in haemostatic testing (32). Some of these factors are dealt with in the following section for INR testing.

Both venous and capillary blood can be used for INR measurements depending on the type of instrument used. For citrated venous blood, it is recommended to use test tubes with 3.2% sodium citrate (33). In most cases, it is not necessary to include a discard tube even if INR is tested on the first tube drawn (33-35). However, if a winged blood collection set ("butterfly") is used and the citrate tube is the first tube, a discard tube is necessary to avoid under filled tubes (33). The citrate tube should ideally be completely filled, meaning that the blood to citrate ratio is 9:1. Under filled tubes increase the citrate concentration and the dilution of the blood sample, meaning that the clotting times and thus the INR value will increase. The recommended fill volume is 90% (33), but volumes as low as 67% (36) and 60% (37) have shown no significant influence on the INR results. It has been reported that inappropriate filling volume is the most common pre-analytical error in INR testing (38).

If the INR measurement is performed in plasma, the citrated whole blood should be centrifuged at 1500g for minimum 15 minutes in order to produce platelet-poor plasma ($<10x10^9$/L) (33). The test tubes should be stored at room temperature (33), both as citrate whole blood and as citrate plasma, as well as empty tubes. Storage in the refrigerator can cause cold activation of factor VII (33) and decreased INR values. A 24 hour stability limit is generally recommended by CLSI for INR samples (33), but sample stability may, however, be dependent on the system used, and each laboratory should therefore confirm stability with its own system (33, 39, 40). A study investigating the stability of a Stago instrument/reagent combination showed that the samples were stable for 52 hours (41). A pilot study was conducted prior to this thesis (unpublished data) where plasma samples were analyzed at day 0 (within 2 h), day 1 (within 24 h), day 2 (within 48 h) and day 3 (within 72 h) after collection. The samples were stored as citrated whole blood at room temperature and mail transport was simulated. The result showed no significant INR changes after 72 hours ($p=0.79$), which is similar to the findings of others (42, 43).

The effect of haemolysis in the samples is not clear (33). However, if haemolysis is present, it is recommended to use an instrument with mechanical clot detection rather than optical clot detection (33, 40). This recommendation is also valid for samples with lipemia and bilirubinemia (33).

1.4.2 Analytical phase, analytical quality

The analytical quality of a laboratory method should be good in order to ensure safe patient treatment. Therefore, it is important to frequently control the analytical quality by internal quality control (IQC) and external quality assessment (EQA). However, the control is useless without knowing how good the quality should be and which aspects of quality that is controlled. Therefore, it is important first to determine (or agree on) the analytical quality specifications, then estimate the analytical performance of the method, and finally assess the analytical quality by suitable control systems (44).

Different aspects of analytical performance characteristic, analytical quality specifications and analytical quality control are described in more detail in the following sections with special emphasis on POC INR methods.

1.4.3 Post-analytical phase

The post-analytical phase is not dealt with in this thesis.

1.5 Analytical performance characteristics

The quality of the analytical performance can be characterized in different ways. It can be expressed qualitatively as “good” or “bad” or quantitatively in numerical terms. The types of analytical errors can be divided into systematic and random errors, which can be added into a total error:

Total error = systematic error + random error

The analytical performance characteristics for total analytical error, systematic analytical error and random analytical error are expressed in qualitative terms as accuracy, trueness and precision, respectively, and in quantitative terms as inaccuracy, bias and imprecision, respectively, as illustrated in Table 1.

Table 1: Analytical performance characteristics (modified from Menditto et al. (45)).

Type of errors	Qualitative term of performance characteristics	Quantitative term of performance characteristics
Total error	Accuracy	Inaccuracy
Systematic error	Trueness	Bias
Random error	Precision	Imprecision

Accuracy is defined as closeness of a single measurement value and the true value (46), and the inaccuracy can be expressed in concentration units or in percentage. The accuracy includes the effects of both trueness and precision.

Trueness is defined as closeness of agreement between the average of replicate measurements and a reference value (46), and the bias can be expressed in concentration units or in percentage. Method bias is the mean deviation of results from one method compared with results from a reference method (or a designated comparison method which is traceable to a reference method). In IQC terminology, this is called the *stable* systematic error and should ideally be corrected before patient results are released (47). In IQC, it is the *changes* in the systematic errors we will detect.

Precision is defined as closeness of agreement between replicate measurements of the same sample under specified conditions (46). The three conditions are, ranged from less to more variable conditions: repeatability, intermediate precision and reproducibility. Repeatability and intermediate precision are estimates of the imprecision in a single laboratory, whereas reproducibility is an estimate of the imprecision between different laboratories (48). The terms "within- and between-laboratory variation" used in this thesis can be referred to as intermediate precision and reproducibility, respectively. The total within-laboratory variation consists of within-run variation and between-run variation. The within-run variation can be referred to as repeatability. The imprecision is called analytical variation and can be expressed numerically in concentration units as standard deviation (SD) or in percentage of the mean value as coefficient of variation (CV). In IQC terminology this is called the *stable* analytical variation, and it is the *increase* in this variation we want to detect.

The term "single participant performance" is used in this thesis to describe the performance of a single participant in EQA (i.e. the performance of the user using a specific POC method).

1.5.1 Special issues regarding INR methods

Trueness

Trueness of INR methods is difficult to assess because of the lack of a functional reference method (49). The top of the INR hierarchy is the conventional reference measurement procedure: the manual tilt-tube technique with the use of a certain reference reagent (reference thromboplastin, IRP). The manual tilt-tube technique cannot be used directly to establish the true value of a patient sample (or an EQA control material), because different IRPs can give different INR values (8, 50, 51). Consequently, the reference method is only used to calibrate other methods. Different methods should be calibrated with different IRPs (see 1.3). No true calibrators have been developed (52). According to the database of the Joint Committee for

Traceability in Laboratory Medicine (JCTLM) (53), there is no reference measurement procedure available for prothrombin time testing, but there is, however, one higher-order reference material available (rabbit brain thromboplastin). It is important to remember that INR results are not traceable to a SI unit; it is not a quantification of a single component, but a measurement of the result of a reaction involving many factors, e.g. a cascade system. Consequently, there are no primary reference measurement procedure and no primary calibrator.

Because of these issues it is common to perform comparisons between routine methods. Harmonization of methods used in primary care and in hospitals could improve the anticoagulation treatment, and it is therefore valuable to assess the agreement between a POC INR method and a hospital routine method that is traceable to a reference measurement procedure (with a given IRP).

Precision

For capillary POC INR methods, such as for example the CoaguChek instruments, it is not possible to estimate ordinary within-run variation (repeatability) because each reagent test strip cannot be re-used. In addition, fresh capillary whole blood cannot be analyzed in replicate because of the limited stability (the coagulation process starts immediately after the blood sampling). However, it has been suggested that the repeatability of such methods can be estimated by performing duplicate measurements using blood from two fingers and using one reagent strip per measurement (52, 54). When several patient samples are analyzed in duplicate, the within-run variation can be estimated as (55):

$$SD_{within-run} = \sqrt{\frac{\sum d^2}{2n}}$$

where d is the differences between duplicates and n is the number of duplicate measurements. This formula is used in paper I to calculate the repeatability. The intermediate imprecision cannot be estimated using capillary samples from patients

because of the limited stability. To verify the intermediate precision stable control materials must be used (52).

1.6 Analytical quality specifications

1.6.1 In general

All laboratory measurements have some degree of measurement uncertainty. The true value is unknown; the measurement result is always only an estimate of the true value. We cannot require that a method has no measurement errors; we have to accept some uncertainty. The question is how much we can tolerate to ensure safe patient treatment and diagnosing. Therefore, we need analytical quality specifications in order to assess if the method has good enough quality in everyday practice. Analytical quality specifications are necessary to evaluate the analytical performance characteristics.

Specifications for analytical quality can be determined on the basis of different strategies in a hierarchical system (56). At the top of the hierarchy is the strategy that quality specifications should be based on clinical outcomes, followed by specifications based on biological variation and on the clinicians' opinions. Next is specifications based on the professional recommendations (national and international), followed by recommendations from EQA organizations. Finally, at the bottom of the hierarchy is specifications based on "state of the art".

1.6.2 Specifications related to internal quality control

A laboratory method is defined as "in control" if the control results are within the quality control limits, and "out of control" if the results are outside these limits (57). The control limits are depended on the control rule used (see 1.7.2). However, before choosing a control rule we must determine the maximum allowable error that is

acceptable to pass the control system. This maximum error should be determined based on strategies as high as possible in the hierarchy.

When a control rule is chosen the laboratory should calculate the actual total error that is accepted to pass the control system ($TE_{control\ system}$). The magnitude of this error is depended on the chosen control rule, the number of control samples, the analytical imprecision (CV_A) and the bias of the method, and can be calculated as (58):

$$TE_{control\ system} = bias + 1.65 \cdot CV_A + k \cdot CV_A$$

The factor *k* is specific for each control rule and can be found in power function diagrams where the probability of error detection is 90% (59) (for the control rules 1_{2S} and 1_{3S} for n=1 control sample, the *k* is 3.35 and 4.35, respectively). $TE_{control\ system}$ should not exceed the quality specification for total error. From the equation above it is clear that if the stable analytical performance (bias and imprecision) of a method is close to the maximum allowable error, there are minimum room for *changes* in the analytical performance. Therefore, Bolann and Aasberg (60) have suggested to include a safety margin in the quality specification for total error, e.g. to include the power of the control rule used. As a consequence, they also suggest to tighten the specification for analytical variation based on biology, e.g. the imprecision should not be larger than 0.15 of the within-subject biological variation.

In the "six sigma" concept, the maximum allowable imprecision and bias (CV_{max} and $bias_{max}$) are determined based on the quality specification for total error (TE_{max}) (57):

$$CV_{max} \leq \frac{TE_{max}}{6} \qquad \text{and} \qquad bias_{max} \leq 1.5 \cdot CV_{max}$$

If both of these specifications are fulfilled, a sigma performance of at least 4.5 is achieved. The sigma performance describes how many sigmas that fit within the limit for total error and can be calculated as:

$$Sigma = \frac{TE_{max} - bias}{CV_A}$$

A sigma performance of six is characterized as “world class”, whereas a sigma performance of three or less is considered unacceptable (57). However, the “six sigma” concept is only a tool to control the analytical quality. The quality specification for total error must be determined based on strategies in the hierarchical system.

1.6.3 Specifications related to external quality assessment

The acceptability limits used in EQA should of course also be established based on strategies as high as possible in the hierarchy. However, the quality specifications established by EQA organizations are mainly based on a compromise between “state of the art” and biological variation (61).

The quality specifications for inaccuracy, bias and imprecision should be used to assess the performance of different methods and the performance of each single laboratory’s method. However, only a few types of EQA schemes are designed to assess all of these performance characteristics (see 1.7.3). Some schemes can only assess the between-laboratory variation within methods.

The quality specifications used in EQA to assess each single participant can be determined by different approaches (62), e.g. a deviation limit in percent from the assigned target value, a deviation limit in concentration units from the assigned target value, or a z-score limit. The z-score is defined as:

$$Z = \frac{x - \text{target}}{SD_{between}}$$

where x is the participant result, *target* is the assigned target value and $SD_{between}$ is the between-participant standard deviation. A z-score less than two is considered acceptable, whereas a z-score between two and three is a warning signal, and above three is an alarm signal (62). The z-score can thus be compared with the control rules 1_{2S} and 1_{3S} used in IQC (see 1.7.2), although the z-score limits are depended on the

analytical quality of all participants whereas the control rule limits are depended on one lab only. The z-score is based on "state of the art". It has been argued that EQA limits based on biological variation is more useful than the z-score (63).

Some EQA organizers calculate the uncertainty of the assigned target value (or use a fixed uncertainty value) before calculating the acceptability limits. This approach results in different acceptability limits in percent in different levels and assures that the evaluation of the participants' performance is not so dependent on the concentration of the control material.

Different graphical presentation of the data with the acceptability limits (e.g. histogram, Youden plots, Shewhart control charts for results over several surveys) can be given to the participants (62). The scores can be given simply as "acceptable" or "unacceptable", or divided into categories such as "excellent", "very good", "borderline", "insufficient", "poor" or "bad".

1.6.4 Analytical quality specifications for INR methods

The analytical quality specifications for INR methods based on different strategies are mentioned in the following section and are listed in Table 2.

Based on clinical outcome

It has been suggested that the assessment of agreement between INR methods should be based on clinical decision-making (64). In this study they have developed a model to assess the clinically important differences between INR methods, in which they have added saw-toothed lines in Bland-Altman plots. Results are defined to be in agreement if both results lead to the same clinical decision. Another approach is to use error-grids to assess clinically relevant INR differences (65, 66). The INR differences were divided into different error zones (risk zones) with increasing clinical relevance; Zone A - clinical irrelevant (no risk), Zone B - low clinical

relevance (low risk), Zone C - moderate clinical relevance (moderate risk) and Zone D – major clinical relevance (high risk, dangerous).

Based on biology

The general concept is that the analytical performance (imprecision and bias) should be a minor part of the biological variation, meaning that it should make only a small contribution to the total variation of a patient result. The specifications for imprecision and bias are based on the within-subject biological variation (CV_{WB}) and the between-subject biological variation (CV_{BB}) (67).

For INR, different estimates of CV_{WB} and CV_{BB} are reported. The CV_{WB} for healthy individuals are reported to be 2.3% (68), 3.0% (69), 3.3% (70) and 5.8% (71). The CV_{BB} are reported to be 4.0% (69), 5.7% (70) and 6.8% (68). In the database of Ricos et al. (72), the CV_{WB} and CV_{BB} for INR are given as 4.0% and 6.8%, respectively (the database is based on two studies (68, 71)). However, it seems to be higher for patients in steady-state anticoagulation treatment, where a CV_{WB} of 9.0% (73), 9.5% (74), 9.1%-10.9% (75), and 10.8%-13.3% (76) have been reported. Based on these estimates of biological variation different analytical quality specifications for INR can be calculated (Table 2). The specification for imprecision (CV_A), bias and total error (TE) in Table 2 is based on the following formulas:

$$CV_A \leq 0.5 \cdot CV_{WB} \qquad bias \leq 0.25 \cdot \sqrt{CV_{WB}^2 + CV_{BB}^2}$$

$$TE = bias + 1.65 \cdot CV_A \qquad \text{(one-sided 95\% confidence interval)}$$

The formula for TE has, however, been questioned (77). The specification for CV_A is determined under the assumption that bias is negligible and the specification for bias is determined under the assumption that the CV_A is negligible. Thus, the imprecision and bias are interrelated (78) and it is questionable whether it is correct to add these two specifications into a total allowable error.

Based on clinicians' opinions

In a study by Kristoffersen et al. (79), the clinicians in general practice in Norway were asked how much the INR value had to change in order to change the warfarin dose (critical difference between two consecutive measurements). The median critical difference was 0.8 INR, which corresponded to a CV_A of 4.1% (when bias is zero). In another study (80), the clinicians in 13 countries were asked the same question and the median critical difference was 1.9 INR, which corresponded to a CV_A of 11.5%.

Based on professional recommendations

The international organization for standardization (ISO) recommends different requirements for the manufacturer and for the patients INR instruments used for self-testing (52) (Table 2). The requirements for the manufacturers are that 90% of the results in level below 2 INR should be within ±0.5 INR of the reference method. In level 2.0-4.5 INR, the requirement is 90% of results within ±30% of the reference method. Bias should not exceed ±0.3 INR in level 2-4.5 INR. The requirements for the individual POC INR instrument used by the patients are that 95% of the results in level 2-4.5 INR should be within ±0.5 INR, and bias should not exceed ±10%. There are no specifications for imprecision. The quality specifications are based on "state of the art" (52).

The Norwegian INR committee has recommended that 95% of the POC INR results should be within ±20% of a hospital method calibrated with the Swedish calibrators (20). The repeatability (CV_A) should be less than 6%.

In a Danish consensus document, a group of experts have recommended that bias and imprecision of hospital INR methods should be less than 3% (81). Bias and imprecision of POC INR methods should be less than 6% and 5%, respectively. A single POC INR EQA result is considered acceptable if it deviates less than 17.7% from the method specific target value, or less than 26.0% from a comparison method (split sample procedure with the use of patient samples).

The British Society for Haematology has recommended that single INR results from two different methods should be within ±0.5 INR of each other (82). This recommendation is valid both when two POC methods are compared (i.e. when one single INR result analyzed by a patient is compared with one single INR result analyzed by a healthcare professional) and when a POC method is compared with a hospital method (i.e. when one single INR result analyzed by a healthcare professional is compared with one single INR result analyzed at the hospital).

Table 2: Analytical quality specifications for INR based on different strategies.

Strategy for establishing analytical quality specifications	**Quality specification for total error (inaccuracy)**	**Quality specification for systematic error (bias)**	**Quality specification for random error (imprecision)**	**Publisher (comments)**
Based on clinical outcome	Results are in agreement if they give the same clinical decision			Shermock et al. (64)
	Error zones with different risks and clinical relevance			Hemkens et al. (65) Petersen et al. (66)
Based on biological variation	TE ≤ |5.3|%	Bias ≤ |2|%	$CV_A \leq 2\%$	Ricos et al. (72) (healthy individuals)
	TE ≤ |3.725|%	Bias ≤ |1.25|%	$CV_A \leq 1.5\%$	Wada et al (69) (healthy individuals)
	TE ≤ |4.37|%	Bias ≤ |1.65|%	$CV_A \leq 1.65\%$	Rudez et al (70) (healthy individuals)
			$CV_A \leq 4.5\%$	Besselaar et al. (73) (patients on OAT)
			$CV_A \leq 4.7\%$	Lassen et al. (74) (patients on OAT)
			$CV_A \leq 4.6\%$	Geest-Daalderop et al. (75) (patients on OAT)
			$CV_A \leq 5.4\%$	Geest-Daalderop et al. (76) (patients on OAT)
Based on clinicians' opinions			$CV_A \leq 4.1\%$	Kristoffersen et al. (79) (median in Norway)
			$CV_A \leq 11.5\%$	Kristoffersen et al. (80) (median in 13 countries)
Based on professional recommendations	Below 2 INR: 90% of results within ±0.5 INR Level 2-4.5 INR: 90% of results within ±30%	Level 2-4.5 INR: Bias ≤ |0.3| INR		ISO 17593 (52) (requirements for the manufacturers of POC instruments)
	Level 2-4.5 INR: 95% of results within ±0.5 INR	Bias ≤ |10|%		ISO 17593 (52) (requirements for each patient's POC instrument)
	95% of results within ±20%		$CV_A \leq 5\%$	Trydal et al. (20) (for POC INR instruments)
	TE ≤ |14.25|% Single EQA result < |17.7|% Single EQA result < |26.0|%	Bias ≤ |6|%	$CV_A \leq 5\%$	Danish consensus doc. (81) (for POC INR instruments) (EQA with lyophilized control materials) (Split sample with patient samples as EQA)
	Result from two methods within ±0.5 INR			Fitzmaurice et al. (82)

1.7 Analytical quality control

Control of the analytical quality can be performed when the analytical quality specifications for different analytical performance characteristics are defined. Analytical quality control can be divided into internal quality control (IQC) and external quality assessment (EQA). The type of control material used is essential in both control systems.

1.7.1 Control materials

Ideally, the matrix of the control material should be identical to patient samples so that they are fully comparable (83). This is often referred as "commutability" of control materials. Control samples are commutable if the relationship between two methods are the same as for native patient samples (83), meaning that the control materials consist of the same patient specific factors that can interfere with the tests. However, commutable materials are difficult to achieve mainly because the materials must be stable over time and it is difficult to vary the levels of the constituents. Therefore, the materials are often lyophilized and different levels of constituents are added or artificially reduced. These processes can result in non-commutability (84). In addition, lyophilized materials must be reconstituted with the proper amount of water, and this pre-analytical factor does not reflect usual sample treatment and can result in a larger between-laboratory variation than for fresh patient samples. Studies have shown that some processed samples are commutable, whereas some are not (85), and each processed control material should therefore be tested for commutability (86, 87). Different control materials have shown both smaller and larger between-laboratory variation compared to native patient samples (88).

The control material values should preferably be in a level in which medical decisions are made and in different concentration levels (normal and abnormal levels) (57, 89). The homogeinety and lot-to-lot variation for control materials must be at an acceptable level.

Control materials for INR methods

The optimal is to use native patient blood, but this is a challenge because of the limited stability of the samples (especially capillary whole blood, which must be analyzed within seconds). Therefore, most commercial control materials for INR methods are lyophilized. An alternative to the lyophilized material is to use fresh frozen plasma from centrifuged citrated whole blood. When stored at -74° C or lower the plasma is stabile for at least two years (33). However, after the plasma is thawed (at 37 °C) it must be analyzed immediately (33). This material is suitable for plasma methods and is used in EQA scheme in Norway and Denmark for hospital methods. Fresh frozen plasma cannot, however, be used in EQA schemes for POC INR methods for several reasons: 1) not all POC INR methods can use plasma, 2) the number of POC participants are often very high and it would require a large amount of material, which would result in a practical challenge of making this material several times a year, 3) the material must be sent frozen and kept frozen until it is analyzed, and not all primary care laboratories have a freezer. In addition, it would be very expensive to send the samples with express mail because of the high number of participants. Another challenge of using fresh frozen citrated plasma is that for capillary whole blood methods the materials must be recalcified before use, and this does not reflect usual sample treatment and may increase the between-laboratory variability. The fresh frozen plasma may not be commutable for these methods, as shown in the Netherlands where small differences between reagent lot numbers were seen when capillary whole blood was used, whereas larger differences were seen when fresh frozen pooled plasmas were used (90, 91).

Different kinds of commercial lyophilized control plasmas are available, such as human or animal plasma, pooled or single plasma, artificially depleted normal plasma or plasma from patients. It is preferable to use plasma from humans (57), and it is likely that lyophilized plasma from patients on oral anticoagulation treatment is more commutable than lyophilized adsorbed normal plasmas in which the coagulation factors are artificially depleted.

No previous studies have investigated which types of control materials that are in use in the different EQA schemes for POC INR methods. However, some EQA organizations in Europe (92-94), Australia (95) and USA (96) use commercial lyophilized control materials. Unfortunately, no lyophilized material is suitable for all methods, due to insufficient compatibility with the POC INR methods (matrix effects). This means that different materials are required for the different POC INR methods and method specific target values must be used (see 1.7.3). Another drawback of using lyophilized materials is the necessity of adding water and calcium chloride (the latter is only for capillary methods, such as CoaguChek). For the different lyophilized control materials used in this thesis, see Materials and methods.

As mentioned, the optimum would be to use native patient samples as control materials, and an EQA scheme where such material is used, is needed (a proposal of such a scheme is presented in paper III). An investigation of which types of control materials that are in use in the different EQA schemes for POC INR methods should be conducted (and this was done in paper II).

For IQC of POC INR methods it is common to use commercial lyophilized control materials and/or split sample procedure with native patient samples (see 1.7.2). However, no studies have investigated the effectiveness of these two approaches in their ability to detect error. Such an investigation was therefore performed in paper I.

1.7.2 Internal quality control

The objective of IQC is to monitor the stability of the analytical quality of a method, detect changes in analytical errors and to estimate the analytical uncertainty in routine patient samples. The prerequisites are that control materials with established target values must be used (established by the manufacturer or by the single laboratory), the stable analytical variation of the method must be known, analytical quality specifications and control rules must be chosen, and the control materials must behave similar to patient samples (60).

The laboratory analyzes one or more control samples over time in order to detect changes in random and systematic error of the method compared to previously results. The IQC results are assessed “on the spot” and the analytical performance can thus be assessed continuously.

Power functions

The basic principle in IQC is to detect analytical errors. It is not possible to detect all errors at any time, but the aim is to detect as many and as early as possible. Ideally, the control system should give a signal or an alarm when a true analytical error occurs, and no false alarms should be given.

The probability of error detection (P_{ED}) is depended on the analytical quality of the method, the control rule used, the number of control samples, and the size of the error (97). Thus, the probability of detecting a large error is higher than detecting a small error. Power function diagrams illustrate the probability of detecting changes in systematic or random errors with different control rules and different number of samples (59), and can for example be created by computer simulations. The size of the error is usually expressed in number of analytical SD in order to make the diagrams valid for all methods. However, the size of the systematic error can also be expressed in percent. Figure 8 illustrates the power of detecting a change in bias expressed in percent for a method with an analytical variation of CV_A = 3% or CV_A = 5%, when the stable analytical bias is zero. The probability of detecting a positive change in bias is similar to the probability of detecting a negative change in bias if the stable bias is zero. This assumption is done in the Westgard’s diagrams (59). However, if the method has a bias that is not corrected, the power function diagram is asymmetrical around zero and the power function diagrams should include both positive and negative changes. This is done for one control rule in this thesis (see 3.1.5).

The probability of false rejection (P_{FR}) is shown in power function diagrams where the error is zero (Figure 8) and is thus an expression of false alarms. P_{FR} is the same

as type 1 error or α-error (reject the null hypothesis when it is true) and should be as low as possible. P_{ED} is the same as 1 minus type 2 error or 1 minus β-error (β-error is to keep the null hypothesis when it is not true), and should be as high as possible. The latter can be described as the power of a statistical test (98).

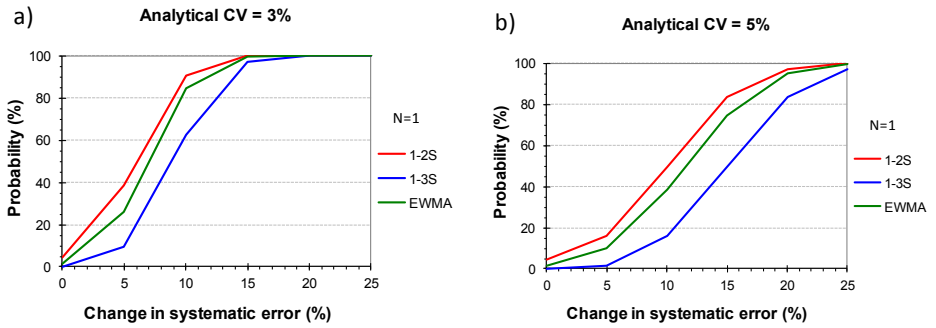

Figure 8: Example of power function diagrams for three different control rules used on a laboratory method with an analytical variation of a) CV = 3% and b) CV = 5%, when the number of control sample is one, and the stable analytical bias is zero. The probability of detecting a change in systematic error of for example 15% for a method with CV_A = 3%, is above 90% for all three control rules, whereas it is decreased to 84% for the 1_{2S} rule, 75% for the EWMA rule, and 50% for the 1_{3S} rule if the method has a CV_A of 5%. The probability that the control system will give an alarm when there is no change in systematic error, is less than 1% for the EWMA and 1_{3S} rule, and 5% for the 1_{2S} rule regardless of the methods analytical variation.

Control rules

Analytical quality specifications must be decided before choosing the control rules because the magnitude of errors that will be detected by the control system is different for the different control rules (57). One should choose a control rule with the lowest possible number of controls, low P_{FR}, high P_{ED}, and the easiest statistical calculations (59). It is recommended that P_{FR} should be less than 5% and that P_{ED} should be higher than 90% (99). This can sometimes be difficult to achieve, and a high P_{ED} should be prioritized if the prevalence of error is high and a low P_{FR} should be prioritized if the prevalence of error is low (100). If the analytical quality of the

method is very stable, a P_{ED} of 50% is acceptable if the P_{FR} is low and the number of samples is low (101).

There are several different control rules, known as the Westgards rules (101), but only the 1_{2S} and the 1_{3s} rule are used in this thesis (paper I). Both these rules depend on the analytical variation (SD_A) of the method, meaning that the limits are numerically wide if the SD_A is large and vice versa. When the 1_{2S} and 1_{3S} rules are used each single control result is assessed on the spot, and only the latest control result is assessed. These rules are thus not suitable to detect changes over time. A trend detection control rule such as the moving average or the exponentially weighted moving average (EWMA) can be used for this purpose. EWMA is the geometric weighted sum of the latest control result and all previous results, in which each single weighted value decreases exponentially with time (102). The formula is shown below (103):

$E_t = \lambda \cdot x_t + (1-\lambda) \cdot E_{t-1}$ general formula

$E_1 = \lambda \cdot x_1 + (1-\lambda) \cdot E_0$ result no. 1

$E_2 = \lambda \cdot x_2 + (1-\lambda) \cdot E_1$ result no. 2

$E_3 = \lambda \cdot x_3 + (1-\lambda) \cdot E_2$ result no. 3

etc

$\mu \pm q \cdot \sigma \cdot \sqrt{\frac{\lambda}{2-\lambda}}$ Upper and lower limit

E_0 is the expected mean value (μ), E_1, E_2, E_3 and so on are the calculated EWMA values, x are each single control result, σ is the stable SD_A, q is a variable that represents the distance from μ expressed in number of standard deviations, and λ is the weighting factor. A high λ weighs the most recent control results more than older results, whereas a low λ weights the older results more than newer results.

A systematic error can arise stepwise or gradually. The source of stepwise changes can for example be calibration differences, change of reagent lot number, and wrong use of pipettes. The reasons for the gradually changes are more diffuse and can be difficult to detect because the systematic error has evolved over time. The EWMA rule gives a relatively quick response to small stepwise and gradually changes, while the 1_{2S} and 1_{3S} rule give a quick response to large stepwise changes.

Internal quality control of POC methods

IQC of POC methods is challenging for many reasons. The persons who operate these methods have usually no or little experience with laboratory work, they lack knowledge of the purpose with IQC, and they have problems with all the statistics (e.g. calculation of the analytical SD, calculation of the control limits). The methods are often easy to use and it is tempting to think that no user errors will occur. In addition, the manufacturers often state that IQC is not necessary because many POC methods have several integrated control systems ("on-board" QC systems) both in the instrument and in each reagent strip, and thus they claim that the method will give an error message if something is wrong with the tests. However, a collective opinion paper states that such controls are insufficient (104). Another challenge is the costs; the tests and control materials can be expensive and some materials have short stability (e.g. the manufacturer's controls for CoaguChek XS Plus must be analyzed within 30 min).

Despite these challenges, it is recommended that IQC of POC INR methods should be performed regularly and preferably every day the instrument is in use (82, 105). In addition, IQC should be event driven, i.e. performed when there is a change in reagent lot, when an unexpected result appear, when there is a change in the measurement procedure, when there is a suspicion of errors, and after instrument maintenance or service (89). The POC instruments should be controlled with control materials or by performing split sample analysis (89). However, it is little evidence whether IQC of POC methods actually leads to quality improvement. IQC is useful

only when an “out of control” result is followed by troubleshooting and corrective actions. Patient results should not be reported until the error has been detected. It is therefore important that the false rejection rate is low so that unnecessary problems and stops can be avoided and that the primary care laboratories can take an analytical control rejection seriously.

Internal quality control of POC INR methods in Norway

Participants in EQA schemes organized by the Norwegian quality improvement of primary care laboratories (Noklus) are regularly asked how often they perform IQC of their method, and the results from the November 2012 survey amongst 1726 participants were as follow: daily 7%, weekly 40%, occasionally 15%, never 2%, when change in reagent lot 15%, and 21% did not answered the question. The Thrombotrack and Simple Simon participants were dominant in the “daily group” and the CoaguChek participants were dominant in the “never group”. This is logical since it is easier to obtain a suitable control material to the first two instruments, which are venous wet chemistry methods, than the latter instruments which are capillary dry chemistry methods (see Material and methods).

Ideally, the control materials used for IQC should be independent of the manufacturer of the method (106). However, the most common approach for IQC for POC INR methods in the Norwegian primary care are that they buy a commercial control material which is made for their method (manufacturer depended). One single control sample measurement is plotted against a target value with defined acceptability limits. These limits can be based on the laboratory’s own analytical variation (usually the Westgard rule 1_{2S} or 1_{3S}) or recommended by experts (e.g. ±10 %). However, the most common is to use the manufacturer’s acceptable target intervals. In addition, or as a substitute to the use of commercial control materials, the primary care laboratories often perform a split sample procedure with native patient samples, e.g. a fresh patient sample is analyzed on their own method and the sample is sent to a local hospital for comparison. The deviation between the two results is then compared with

some defined acceptability limits (e.g. ±20 %). Split sample procedure has been recommended as an alternative when the commercial control materials are not suitable (93, 105), but no studies have compared the utility of these two procedures, i.e. how good the split sample procedure and the use of control materials are in detecting errors.

1.7.3 External quality assessment

One of the main objectives in EQA is to compare results within and between methods. The optimum is to use a reference method in order to assess the trueness of methods. EQA is both a tool to compare the analytical performance of different methods and a tool for the individual laboratory to compare results with others with the same or different methods. EQA is administered by an external part (an EQA organizer) and the result is assessed retrospectively. The time period between the reported result and the feedback should be as short as possible.

The different types of EQA schemes can be divided into six categories (107) and different aspects of analytical quality can be assessed (Table 3). Category 1 and 2 are characterized by the use of commutable control materials with reference target values, in which the trueness of methods can be assessed. In addition, standardization and harmonization of methods can be evaluated by calculating the systematic deviation between methods and the between-method variation. The between-laboratory variation within and between methods can also be evaluated. The trueness of a single laboratory is assessed when mean of replicate measurements is compared with a reference target value (category 1), whereas the accuracy is assessed when a single measurement is compared with a reference target value (category 2). In addition, the single laboratory can compare its own result with the median of other methods and the median of own method. In category 3 and 4 there are no reference target value, and the traceability of methods cannot be evaluated. However, the systematic deviation between methods can be assessed because commutable control samples are used, and thus the between-method variation can be evaluated. In

addition, the between-laboratory between-method variation and the between-laboratory within-method variation can be calculated. The single laboratory can compare results within own method and between other methods. Category 5 and 6 are characterized by the use of non-commutable control samples and there is no reference target value. In these schemes, method specific target values are the only option, and assessment of traceability and harmonization of methods is not possible. However, the between-laboratory variation within methods can be evaluated, and for category 6 this is the only aspect of analytical quality that can be assessed. The single laboratory can only compare results within its own method. In schemes where replicate measurements are performed (category 1, 3 and 5) the mean within-laboratory variation and the within-laboratory variation for a single laboratory can also be assessed.

Table 3: Different EQA categories and the aspects of analytical quality assessment (modified from Miller et al. (107)).

EQA category	Commutable material	Reference target value	Replicate measurements	Assessment of the analytical quality of methods	Assessment of the single laboratory
1	Yes	Yes	Yes	Bias Systematic deviation between methods Between-laboratory variation between and within methods Within-laboratory variation	Bias Systematic deviation from other methods Systematic deviation from own method Within-laboratory variation
2	Yes	Yes	No	Bias Systematic deviation between methods Between-laboratory variation between and within methods	Inaccuracy Deviation from other methods Deviation from own method
3	Yes	No	Yes	Systematic deviation between methods Between-laboratory variation between and within methods Within-laboratory variation	Systematic deviation from other methods Systematic deviation from own method Within-laboratory variation
4	Yes	No	No	Systematic deviation between methods Between-laboratory variation between and within methods	Deviation from other methods Deviation from own method
5	No	No	Yes	Between-laboratory variation within methods Within-laboratory variation	Systematic deviation from own method Within-laboratory variation
6	No	No	No	Between-laboratory variation within methods	Deviation from own method

External quality assessment of POC INR methods

It is recommended that users of POC INR methods should participate in an EQA scheme whenever available (89, 108). However, it is a challenge for the EQA organizers to provide such schemes because of the often large number of participants and that the different methods cannot use the same control materials. In addition, it is a challenge to obtain suitable control materials.

When lyophilized control materials are used in EQA schemes for POC INR methods the schemes are characterized as category 5 or 6. Different materials are used for the different POC INR methods and method specific target values are necessary. Consequently, the assessment of trueness and harmonization of methods is not possible. For category 6, which is most common for POC INR methods, only the between-laboratory variation within methods can be assessed, and the single laboratory can only compare results with its own method. Alternative approaches have therefore been developed but all have some limitations. The European concerted action on anticoagulation (ECAA) has developed a set of five lyophilized QC plasmas with certified INR values (109) in order to evaluate the trueness of POC INR methods. The drawback of this approach is that the QC set is aimed at one type of instruments only (the CoaguChek monitors), and that lyophilized plasmas are used, i.e. it does not necessarily reflect the quality of native samples (91, 110). In addition, the participants cannot compare results with other participants, and this approach is therefore not an EQA but more similar to control systems in which the methods are controlled with certified reference materials.

Another alternative EQA approach is the split sample procedure in which native patient samples are used to compare single INR-results between a POC method and a hospital method (93). The main drawback of this approach is that it is difficult for the participants to find the origin of a *single* deviant result because of the numerous of variables that can affect the result. This split sample approach has also been used to compare results between patient self-testing POC instruments both with POC instruments used in primary care and with hospital methods (111, 112).

There is a need to develop an EQA scheme in which commutable control materials are used and the trueness of all POC INR methods can be assessed. The participants should be informed of the origin of a deviant result (is it due to the method or the participant performance?), and the scheme should be easy for the EQA organizer to perform on a regular basis.

2. AIMS

The main objective of this thesis was to evaluate and suggest improvements of the analytical quality control of INR methods used in primary care.

Paper I (internal quality control)

The aim of paper I was to evaluate and compare two common approaches of IQC of POC INR methods in their ability to detect changes in systematic errors. The two approaches were the use of lyophilized control materials and the use of patient split samples, both in combination with different control rules.

Paper II (external quality assessment)

The aim of paper II was to investigate if and how the European countries provide EQA schemes for POC INR testing.

Paper III (external quality assessment)

The aim of paper III was to develop an EQA model for POC methods that will improve EQA schemes in situations where non-commutable control materials are used.

3. MATERIALS AND METHODS

3.1 Paper I and III:

The use of patient samples and lyophilized control materials in internal and external quality control of INR methods in primary care

3.1.1 Selection of primary care laboratories and patients

The primary care laboratories were randomly selected in paper I, because they should represent a random primary care laboratory performing IQC of their POC INR instrument. In paper III, the primary care laboratories were specifically selected based on certain selection criteria (e.g. good analytical quality and skilled personnel). This was done because these laboratories should establish the systematic deviation (bias) of one type of POC INR method and thus the user errors had to be minimized (these selected primary care laboratories were considered as experts). All primary care laboratories in both studies received written information about the study and gave written informed consent to participate.

In both papers, each primary care laboratory selected patients on long-term oral anticoagulation treatment with warfarin, who came for a routine follow-up visit with their general practitioners. The patients received written and oral information about the studies and gave oral informed consent to participate. The *Regional Committee for Medical and Health Research Ethics* in Western Norway approved both studies.

The number of selected primary care laboratories and patients in the two studies are shown in Table 4. Eighteen primary care laboratories participated in the study in paper I, whereas 72 and 69 participated in the first and second survey in paper III. Four or five patient samples were analyzed at each laboratory. The aim was to get minimum 20 results for each POC method in paper I and minimum 100 results in

paper III. The patient results in paper I were used as input in the computer simulation program whereas the patient samples in paper III were used to estimate method bias.

Table 4: Number of selected primary care laboratories and patients in paper I and III.

POC INR instrument	Number of primary care laboratories	Number of patient samples per lab	Total number of patient results	Paper
Thrombotrack	8	5	40	I
CoaguChek S	5	5	25	I
Hemochron Jr. Signature	5	5	25	I
Thrombotrack	25 and 23	4 to 5	102 and 103	III
CoaguChek XS Plus	23 and 24	4 to 5	99 and 105	III
Simple Simon	24 and 22	4 to 5	113 and 100	III

3.1.2 Collection of split sample results with native patient samples

The principle of the split sample procedure is that a fresh patient sample is analyzed both on the POC method at the primary care laboratory and on a designated comparison method in order to compare results between methods. Single split sample comparisons were evaluated in paper I (used as internal quality control), whereas mean of approximately 100 split sample comparisons were evaluated in paper III (used as method comparison). In paper I, the selected primary care laboratories analyzed an INR split sample once a week for five weeks. The mean deviation of all split sample results and the within-day and between-day variation were used as basis (input) for the computer simulations. In paper III, each selected primary care laboratory analyzed four or five split samples during a two weeks period, and mean deviation of all split sample comparisons for one POC INR method were calculated. This estimation of POC method bias was then evaluated against an analytical quality specification of 6%. All patient samples in paper I and III were sent to the hospital laboratory (Haraldsplass Diaconess Hospital, Bergen, Norway) by ordinary mail the same day of sample collection, and the samples were analyzed the day of arrival.

3.1.3 Collection of control results with lyophilized materials

In paper I, the selected primary care laboratories analyzed a commercial lyophilized control material once a week for five weeks. The control samples were analyzed at the same day as the split samples. The mean of all the control results and the within-day and between-day variation were used as basis (input) for the computer simulations. In paper III, results from two ordinary EQA schemes for POC INR provided by Noklus were used. The participants analyzed two lyophilized control materials in each scheme. The selected expert primary care laboratories were amongst these participants, and the target values were established based on the results from these experts. All participants were evaluated against these target values with the acceptability limit of ±15% from the target value. The different control materials used in paper I and III are shown in Table 5.

Table 5: Lyophilized control materials used in paper I and III.

Name of control	Material	Reconstitution	Manufacturer	Used on POC INR instrument	Paper
CoaguChek PT Controls	Lyophilized rabbit plasma	Breakable ampoule with water and calcium chloride are integrated	Roche Diagnostics, Germany	CoaguChek S	I
DirectCHEK, abnormal	Lyophilized whole blood from sheep, horse and bovine	Breakable ampoule with water are integrated	International Technidyne Corporation (ITC), USA	Hemochron Jr. Signature	I
Control Plasma AK	Lyophilized patient plasma	Separate vial with water to be added	Baxter AG, Austria	Thrombotrack	I and III
TriniCHEK	Lyophilized adsorbed human plasma	Separate vial with water to be added	Trinity Biotech, Ireland	Thrombotrack and Simple Simon	III
OKP	Lyophilized bovine and adsorbed human plasma	Separate vial with water to be added	MediRox AB, Sweden	Simple Simon	III
NEQAS Level 1 and 2	Lyophilized adsorbed human plasma	Separate vials with water and calcium chloride to be added	UK NEQAS for Blood Coagulation, United Kingdom	CoaguChek XS Plus	III

3.1.4 INR methods

The INR methods used in paper I and III are shown in Table 6. At the time, this was the most commonly used INR methods in Norway. A total of five different POC INR instruments were used and the designated comparison methods were the STA Compact and the STA R Evolution instruments with the SPA reagent. All of the INR methods differ in one way or the other (Table 6), and all have different traceability (Figure 6). The two comparison methods are, however, almost identical.

Table 6: INR methods used in paper I and III.

Instrument	Type of method	Clot detection	Thromboplastin	Chemistry (dilution)	Paper
CoaguChek S	Quick	Optical	Rabbit	Dry (none)	I
Hemochron Jr. Signature	Quick	Mechanical	Recombinant human	Dry (none)	I
STA Compact	Owren	Mechanical	Rabbit	Wet (1:21)	I
Thrombotrack	Owren	Mechanical	Bovine	Wet (1:5)	I and III
Simple Simon	Owren	Optical	Rabbit	Wet (1:21)	III
CoaguChek XS Plus	Quick	Electrochemical	Recombinant human	Dry (none)	III
STA R Evolution	Owren	Mechanical	Rabbit	Wet (1:21)	III

3.1.5 Computer simulations

In conjunction with paper I, we designed a computer simulation program in Microsoft Excel 2003. The program is a tool for evaluation of different control system regarding the ability of detecting systematic errors and the rate of false alarms. Two simulation programs were developed, one for the lyophilized control sample results and one for the patient split sample results. Four different control rules were used in each program. The 1_{2S}, 1_{3S} and EWMA rule were used in both programs, whereas the deviation limits of ±10% and ±20% (called $1_{10\%}$ and $1_{20\%}$) were used in the control sample simulation and the split sample simulation program, respectively. The control rules were selected because they are commonly used in primary care, except for the

EWMA rule which was chosen because it is a trend detection rule. The stable bias was ignored for the 1_{2S}, 1_{3S}, EWMA and $1_{10\%}$ rule, but was not ignored for the $1_{20\%}$ control rule. Changes in systematic errors of ± 0%, 5%, 10%, 15%, 20% and 25% were simulated in each model.

In real experiments the analytical quality is estimated based on the single measurements. In computer simulations it is the opposite situation; the single measurements are estimated (simulated) based on the analytical quality. Thus, if the analytical quality is known, a large number of single measurements can be simulated without doing large real experiments which can be time consuming and expensive. In paper I, the analytical quality was estimated based on empirical data, i.e. each primary care laboratory analyzed one split sample and one control sample once a week for five weeks, and this was done only once. The simulation programs were designed to simulate that one split sample and one control sample were analyzed once a week for 20 weeks, and this was done 1000 times. However, paper I describes results for one (n=1) and four (n=4) weeks, only. The empirical input data were mean of all control results or mean deviation of all split sample results, the within-laboratory within-day variation and the within-laboratory between-day variation. The total within-laboratory variation was used in the simulation programs to calculate the control limits, and can be expressed as

$$CV_{total} = \sqrt{CV^2_{within-day} + CV^2_{between.day}}$$

The within-day variation can be referred as repeatability, whereas the total within-laboratory variation can be expressed as intermediate imprecision. As stated in section 1.5.1, patient samples cannot be used to estimate the intermediate imprecision because of the limited stability. It is therefore important to notice that the intermediate imprecision was not calculated directly from the patient samples, but from the patient *split* samples (e.g. the differences between the POC method and the comparison method). The $CV_{within-day}$ for the split samples are thus a combination of the repeatability of the POC method and the repeatability of the comparison method.

3.1.6 Quality specifications

In paper I, a change in systematic error of ±15% was considered clinically important and all control systems should be able to detect this error as fast as possible (low number of control samples), regardless of which instruments and control rules that are used and regardless of control materials or split samples are used.

In paper III, the quality specifications were set to ±6% for bias (81) and ±15% for a deviating EQA result. This bias specification was chosen because the specification based on biology is calculated from healthy individuals and is thus too tight, and the ISO bias specification is valid in the therapeutic level or for patient self-testing only (Table 2). The quality specification for a deviating EQA result was chosen because paper II showed that this was the most commonly used acceptability limit for POC INR methods used by EQA organizers.

3.1.7 Statistical calculations

All statistical calculations were performed using Microsoft Office Excel 2003. In paper I, the analysis of variance (AVOVA) nested design (113) was used to estimate the within-lab within-day variation and the within-lab between-day variation, as recommended by ISO (52). In both paper I and III, outliers were excluded according to Burnett (114). The box plots in paper III (supplemental files) were created in SPSS.

3.2 Paper II

Survey amongst EQA organizers in Europe

A total of 30 European countries were invited to participate in this study. The invitation was sent by e-mail to all members of the European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM). To reach as many European countries as possible, non-members were also contacted. In the invitation, which was sent in 2010, they were asked whether or not an EQA scheme for POC INR was provided in their country at that time. Those EQA organizers that reported that they did provide this service were asked to fill in a questionnaire of 23 questions dealing with how the schemes were organized and how they interact with the participants (see Appendix).

4. RESULTS

4.1 Paper I

Internal quality control of INR methods in primary care

The estimated analytical quality of Thrombotrack, CoaguChek S and Hemochron Jr. Signature is shown in Table 7 for the lyophilized control materials and in Table 8 for the patient split samples. These estimates were used in the computer simulation programs to create power function diagrams. In general, the imprecision was lower for the control materials than for the split samples, meaning that the probability of detecting a systematic error was higher for the control samples than for the split samples. This was true for all control rules and all instruments. The P_{FR}, however, was similar for the control samples and the split samples. An example of this is shown in Table 9, where the change in systematic error was set to +15% and the number of observations were set to four. The probability of detecting a change in systematic error of +15%, for e.g. the 1_{3S} rule using four observations on the Thrombotrack instrument, was >99% for the control samples and 62% for the split samples, whereas the P_{FR} were 1% in both control systems (Table 9). Overall, the EWMA control rule had the best combination of high P_{ED} and low P_{FR} for all three instruments.

Table 7: The empirical data for the control samples.

POC instrument	n	Mean INR	CV_{within} %	$CV_{between}$ %	CV_{total} %
Thrombotrack	39	2.19	1.9	3.3	3.8
CoaguChek S	24	2.76	7.3	5.1	8.9
Hemochron Jr. Signature	25	4.73	9.4	0.0	9.4

Table 8: The empirical data for the patient split samples.

Instrument	n	Bias %	CV_{within} %	$CV_{between}$ %	CV_{total} %
Thrombotrack	39	13.5	2.3	6.5	6.9
CoaguChek S	25	7.7	5.9	8.8	10.5
Hemochron Jr. Signature	25	16.7	9.0	11.7	14.8

Table 9: The probability of error detection (P_{ED}) and false rejection (P_{FR}) for control samples and split samples for different control rules and different POC INR instruments in detecting a change in systematic error of +15% when the number of observations were n=4.

Control rule, POC instrument	Control samples		Split samples	
	P_{ED} (%)	P_{FR} (%)	P_{ED} (%)	P_{FR} (%)
1_{2S}				
Thrombotrack	>99	18	97	18
CoaguChek	85	16	74	18
Hemochron	82	19	53	17
1_{3S}				
Thrombotrack	>99	1	62	1
CoaguChek	35	1	22	1
Hemochron	28	<1	9	1
EWMA				
Thrombotrack	>99	2	94	2
CoaguChek	74	1	56	2
Hemochron	67	1	26	2
$1_{10\%}$				
Thrombotrack	>99	3		
CoaguChek	99	72		
Hemochron	99	75		
$1_{20\%}$				
Thrombotrack			>99	54
CoaguChek			97	43
Hemochron			>99	88

4.2 Paper II

External quality assessment of point-of-care INR testing in Europe

A total of 30 European countries were asked whether or not they provide an EQA scheme for POC INR, and 28 responded (Estonia and Lithuania did not respond). Most of the countries (n=19) did not provide this service (Table 10). All 12 organizations in the 9 countries that did provide EQA for POC INR (Table 11) answered the questionnaire.

Table 10: European countries that *did not* provide EQA for POC INR.

European countries that did not provide EQA for POC INR			
Belgium	Iceland	Poland	Slovenia
Bulgaria	Ireland	Portugal	Spain
Croatia	Italy	Romania	Sweden
France	Latvia	Russia	Turkey
Germany	Luxemburg	Slovakia	

Table 11: European organizations that provided EQA for POC INR.

European country	**EQA organization**	
Austria	ÖQUASTA	Austrian Society of Quality Assurance and Standardization
Czech Republic	SEKK	External quality assessment system for clinical laboratories
Demanrk	DEKS	Danish Institute for External Quality Assurance for Laboratories in Health Care
Finland	Labquality	Labquality
Hungary	QualiCont	In Vitro Diagnostic Quality Control Nonprofit Public Utility Ltd.
Netherlands	ECAT and FNT	External quality Control for Assays and Tests Federation of Netherlands Thrombosis services
Norway	Noklus	Norwegian Quality Improvement of Primary Care Laboratories
Switzerland	CSCQ and MQ	The Quality Control Center Switzerland Association of Medical Quality Control
United Kingdom	UKNEQAS and WEQAS	United Kingdom National External Quality Assessment Scheme Welsh External Quality Assessment Schemes

A vide variation were seen between the different EQA schemes. The number of samples distributed per year varied from 1 to 12 with the median of 4, and the number of participants varied from 75 to 2665. All EQA organizations offered schemes to CoaguChek users, and this was the most commonly used instrument in all countries. Different kinds of control materials were used: lyophilized plasma (most common), lyophilized whole blood, frozen artificial liquid material, fresh frozen patient plasma and fresh capillary patient blood. Peer group target values were most common in all schemes. The acceptability limits varied between 12% and 30%, with a limit of 15% as the most common.

4.3 Paper III

New model for external quality assessment of point-of-care testing

In this paper, a new model for EQA was developed (Figure 9). Fresh patient samples were used in combination with lyophilized control materials in order to assess both the method bias and the single-participant performance. The model was applied twice (in 2010 and 2011) on POC INR methods.

The INR method biases estimated from the native patient samples were -0.4% and -1.2% for CoaguChek, 1.1% and 1.8% for Simple Simon, and 9.0% and 8.2% for Thrombotrack, for the 2010 and the 2011 sampling period, respectively. CoaguChek and Simple Simon fulfilled the quality specification of bias less than 6%, while Thrombotrack exceeded this limit.

The portion of participants with INR results within the quality specification of ±15% for a deviating EQA result using lyophilized control materials was between 97% and 98% for CoaguCheck and Thrombotrack users, and between 90% and 97% for Simple Simon users for both control materials in the two EQA surveys.

The best combined assessment scenario was when both results were within the quality specifications. More than 90% of the CoaguChek and Simple Simon

participants achieved the best case scenario result, while none of the Thrombotrack participants achieved this result because of the unacceptable bias (see supplemental files of paper III).

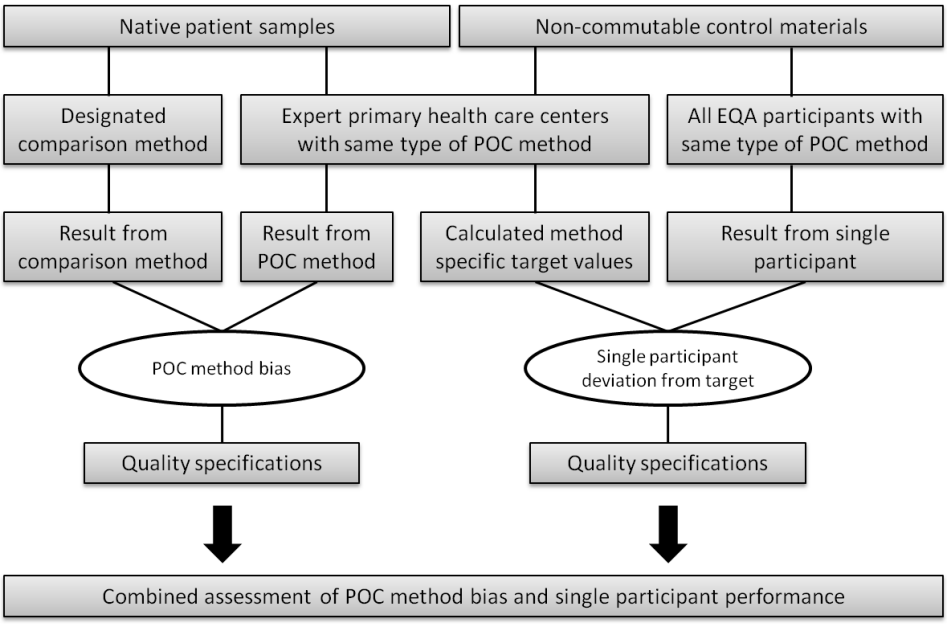

Figure 9: Principle of the new EQA model. Native patient samples should be analyzed both with a reference method or a designated comparison method, and at some selected expert primary care laboratories. The "true" value should be established with the reference or comparison method. Bias of the POC method can then be estimated. In the same time period, non-commutable control materials should be distributed to all EQA participants and to the expert primary care laboratories. A method specific target value for the non-commutable control materials for each POC method should be established based on the results from the experts. Each single participant result can then be compared with this target value. By using separate quality specifications, combined assessment of method bias and single participant performance is possible. The best case scenario is when both results are within the quality specifications. The worst case scenario is when both results are outside the quality specifications and the results deviate in the same direction. Such a participant will report patient results that are greatly over- or under estimated. The idea is to accumulate the results from year to year, so that possible problems better can be addressed: "Is it the method or is it the participant performance that needs to be improved?"

5. Discussion

Patients on warfarin treatment are commonly monitored in primary care. It is vital that the INR methods used in the monitoring process have good analytical quality in order to ensure safe patient treatment. As a consequence, the primary care laboratories are recommended to perform internal analytical quality control and participate in external analytical quality assessment schemes. This thesis and its associated papers aim to evaluate and improve these analytical quality control systems. In the following, some methodological considerations and main findings in the three papers will be discussed followed by some recommendations at the end.

5.1 Methodological considerations

5.1.1 Study population

The primary care laboratories were randomly selected in paper I because they should represent all laboratories in primary care. However, due to the relative small number of participants it cannot be excluded that if other laboratories had participated different estimates of the analytical quality could have been obtained. Nevertheless, the aim of this study was not to evaluate the analytical quality of POC methods but to use these estimates in computer simulations to evaluate different IQC procedures. It is no reason to believe that the main conclusions would have been different if other primary care laboratories had participated. With a simulation of one thousand control runs, the internal validity of this study is considered strong.

In paper III, the primary care laboratories were selected based on certain criteria in order to characterize them as “experts”. This was done because the results from these laboratories were used to estimate method bias and thus the user errors had to be minimized. However, it cannot be excluded that some of the deviating results between the expert’s POC method and the comparison method were due to user error. By choosing other experts the method bias could have been slightly different, but

since the biases were estimated based on a large number (one hundred) of patient samples the external validity is considered strong. In addition, the expert laboratories were selected from all parts of Norway and thus represented all "good" primary care laboratories in Norway.

In paper II, the recruitment of the participants is essential for the results of the study and thus the internal validity. An invitation was sent to 30 of the 44 European countries defined by the United Nation (115). It cannot be excluded that some of the non-invited countries do provide EQA schemes for POC INR, although this is not likely (the non-invited countries were Albania, Andorra, Belarus, Bosnia and Herzegovina, Greece, Liechtenstein, Macedonia, Malta, Moldova, Monaco, Montenegro, San Marino, Serbia and Ukraine). In addition, even if an invited country reported that they did not provide EQA for POC INR, it cannot be excluded that other kinds of quality control for POC INR are provided in that country (e.g. control of patient self-testing performance). This limits the external validity of this study.

5.1.2 Questionnaire

In paper II, a questionnaire was distributed to the EQA organizers that provided EQA schemes for POC INR methods at the time this study was conducted. All organizers answered the questionnaire which consisted of 23 questions, both open and closed (see Appendix). Eleven of the twelve EQA organizers answered all 23 questions; only one organizer did not answer three questions regarding supervision and guidance of participants. This high response rate makes the internal validity of this study strong. The questionnaire was designed to handle several different POC INR schemes for each EQA organizer, because different control materials are often circulated to different POC methods, and different EQA approaches are therefore possible. This increased the external validity of this study.

Since most of the European countries did not provide EQA for POC INR methods it would have been interesting to know the reason why. This question could thus have

been included in the study. However, there are reasons to believe that this is due to the challenge of obtaining suitable control materials, as reported from Germany (116).

5.1.3 Designated comparison method

The choice of comparison method was based on the fact that there is no reference method available for INR and thus a routine laboratory method was chosen. Because patients often are treated both in primary care and in hospitals, is it important to examine the agreement between methods used in these settings. Therefore, the most commonly used hospital method in Norway was chosen as the comparison method in both paper I and III. The designated comparison method was calibrated at the start of each study with the Swedish calibrators using the "direct INR calibration" procedure (18). As an alternative, the designated comparison method could have been calibrated with a reference thromboplastin of "higher order", i.e. with a rabbit IRP using the "full WHO ISI calibration" procedure, in order to give more "true values". This could have increased the validity of the bias estimations in paper III, but it would not have reflected the real situation in Norway where this calibration procedure is not used. The comparison between POC methods and routine hospital methods in paper III could therefore not have been done. In paper I, in which the IQC routines were examined, a "full WHO ISI calibration" procedure would not have been of any value because the aim of this study was to detect *changes* in systematic errors and thus the stable biases of the POC methods were ignored. However, the biases were not ignored for the $1_{20\%}$ rule, but again this calibration would not have represented the common IQC practice in Norway and thus the results would not have been valid.

The designated comparison method is a key factor in bias estimation. ISO (52) state that a routine measurement procedure that is traceable to the manual tilt-tube technique with an IRP may be used to assign the reference values. By choosing a different comparison method in paper III other bias estimates could have been obtained due to the limited harmonization of INR methods (117). To ensure that the

designated comparison method represented “a typical Norwegian hospital method”, the results in paper III were adjusted by using a regression equation based on three fresh frozen control plasmas with assigned values. The assigned values were the median of Norwegian laboratories using the same method as the designated comparison method. The external validity of the estimated biases is therefore good. However, the principles of the proposed model are more important in this paper than the actual results. If a different comparison method had been chosen the principles of the model would still have been valid.

A strength in both paper I and III was that only one lot number of reagents was used on the comparison method and only one lot of lyophilized control materials was used on the POC methods. This avoids possible lot-to-lot variability and strengthens the internal validity of the studies. Different lot numbers of POC reagents were used in paper III, making lot-to-lot comparisons possible, both by using native patient samples and by using lyophilized control materials. This is a strength of this study. In paper I, however, different lots of POC reagents would have resulted in false high between-laboratory variation estimations and thus the computer simulation inputs would not have been correct. Therefore, only one lot of POC reagent was used in paper I.

5.1.4 Split sample procedure

The split sample procedure with native patient samples is commonly used in method comparison studies, both when two POC methods are compared and when a POC method is compared with a hospital method. This procedure has also been used in EQA and IQC. The split sample results are assessed differently dependent on the purpose of the comparison, and this will be addressed in the following sections.

In method comparisons, all split sample results are assessed together and different statistical calculations can be made, such as for example mean deviation (bias) between the POC method and the comparison method, regression analysis and

difference plot (118). The use of split sample procedure in method comparison studies is recommended by the CLSI and ISO (52, 118), and this procedure was modified and used in paper III. The advantage of using split sample is the use of commutable control samples, i.e. fresh native patient samples.

When split sample procedure is used in EQA schemes, one split sample result for each participant is assessed. This approach has been reported useful when there are no suitable control materials available (93). Another study (111) which has compared four different EQA approaches for patients (traditional EQA schemes with and without supervision, split sample procedure between the patient's POC method and the clinic's POC method, and split sample procedure between the patient's POC method and a hospital method) concluded that the best approach was the traditional EQA schemes without supervision (personal guidance) or the split sample procedure between a patient's POC method and a clinic's POC method. The EQA split sample procedure against a hospital method is considered less useful because it is difficult to locate a possible error of the POC method when only one split sample result is assessed (111, 119). This is because it includes pre-analytical factors and the performance of the hospital method. Assessment of a single split sample result was therefore not used in paper III.

The purpose of IQC is to detect changes in analytical quality (random and systematic errors) and to estimate the analytical imprecision of a method. However, when the split sample procedure is used none of these issues are suitable addressed. In paper I, the probability of error detection was higher using a lyophilized control material than using the split sample procedure. This was because the imprecision of the control sample measurements was lower than the imprecision of the split sample measurements. Thus, it was recommended not to use the split sample procedure as a substitute for the use of control materials in IQC. Monitoring of the analytical performance of a single laboratory should be performed by using suitable control materials. The optimum is to use control materials commutable with patient samples. If the materials are non-commutable, the target value must be changed when there is a change in lot number of either the control material or the POC reagent (120). The

split sample results can only be evaluated retrospectively, whereas the results from internal control materials can be evaluated "on the spot". Another drawback of using the split sample procedure in IQC is that the stable analytical bias must be corrected in order to use some of the control rules (e.g. 1_{2S} and 1_{3S}), and this may be impractical and difficult for the primary care laboratories to perform. Consequently, a percent deviation limit is used instead, such as the $1_{20\%}$ rule which is commonly used in Norway. The biases of the POC methods are not corrected, and as shown in paper I this resulted in a high false alarm rate for POC methods with a positive bias in detecting a positive change in systematic error (and vice versa; for POC methods with a negative bias in detecting negative changes in systematic errors). The split sample procedure is thus only suitable for method comparisons, in which one primary care laboratory can estimate the bias of their POC method compared to a hospital method. Estimation of the analytical imprecision of the POC method is not possible using the split sample procedure. This must be estimated either by using a control material or by using patient samples analyzed in duplicate on the POC method.

5.2 Discussion of results

5.2.1 EQA schemes and control materials for POC INR methods

Paper II showed that only a few countries in Europe provide EQA schemes for POC INR methods and that there is a vide variation in how these schemes are organized. It seems that it is a difficult task to obtain suitable control materials, especially for capillary methods. Consequently, alternative variants have been developed. One approach is the split sample procedure with native patient samples, and another is the use of the "on-board" QC system for the INRatio instruments. The latter approach is provided in Austria and the Czech Republic and is developed by the fact that there are no commercial control materials available for this instrument; only fresh patient samples can be used. All participants in the two countries receive test strips from one batch and are asked to analyze one arbitrary patient sample (fresh capillary sample).

The instrument displays two QC values (in seconds) for each patient result and the participants are asked to report the two QC values only. The two target values are the median from all participants from that specific batch of test strip. This EQA approach can be classified as category 4 (see 1.7.3) in which commutable samples are used and evaluation of the participant performance and the between-laboratory variation of one reagent lot is possible. However, evaluation of lot-to-lot variation is not possible since only one lot is distributed to all participants. In addition, the harmonization of methods cannot be assessed because this EQA approach is addressed to one POC method only.

Paper II showed that the most commonly used control material in EQA schemes for POC INR in Europe is lyophilized plasma. Different control materials are provided to the different POC INR instruments, and method specific target values are used. Only one EQA organization (the Federation of Netherlands Thrombosis services) uses a native-like material, i.e. fresh frozen plasma to the CoaguChek participants and fresh frozen plasma with added red blood cells to the ProTime participants (all the participants must in addition add a calcium chloride solution to the materials). These materials are, however, not suitable in EQA schemes for primary care laboratories mainly because of the limited stability (all participants must analyze the controls on the same day). Consequently, these materials are distributed to anti-coagulation clinics only.

Another alternative control material to the lyophilized materials is provided in Wales, where an artificial liquid material is distributed to the CoaguChek participants. The material is based on bovine plasma with added stabilizers and dyes to make it look like whole blood. The advantage of using this material is that neither addition of water nor calcium chloride is required, and the pre-analytical errors are therefore reduced. The drawback is that the material is sent frozen which is a practical and economical challenge for the EQA organizers and that the material is not based on human plasma which may result in non-commutability with patient samples.

5.2.2 Novel EQA model

In paper III, a novel EQA model was proposed. The model is a combination of method comparison and traditional EQA of a single participant. Split sample procedure with native patient samples was used to assess agreement between methods, and lyophilized control plasmas were used to assess single participant performance. The two assessments were then combined in one figure. The POC methods were assessed in the hands of the user. A similar approach has been proposed by Tripodi et al. (112) for POC INR methods used for patient self-testing. The patients analyzed three control plasmas on their POC method and the results were compared with the method specific target value. In addition, a split sample from each patient was analyzed both on the patient's POC method and at a hospital method. The main difference between this study and the proposed model in paper III is that each split sample result was assessed in the Tripodi study (like an IQC result), whereas the mean of all split sample results were assessed in paper III (method bias). Single split sample results are influenced by both bias and imprecision, whereas the influence of imprecision are reduced considerable when mean of all split sample results is assessed. The proposed model in paper III is thus more suitable to assess trueness of POC methods. Another approach for trueness assessment is established by the ECAA where five lyophilized control plasmas with certified INR values are used to assess one type of POC INR method (109). As discussed previously (see 1.7.3), this approach has however some disadvantages. A third approach is developed by Ross et al. (121), in which the matrix effect of a control material is estimated. By correcting for this matrix related bias in the establishment of the reference target value, trueness of methods as well as trueness of each single participant was assessed. This approach is, however, difficult to implement for INR methods because of the lack of a functional reference method (49).

Already in 1996, Libeer et al. (122) suggested that a selected group of participants could use the split sample procedure with native patient samples to estimate method bias once a year. The proposed model in paper III is an elaboration of this suggestion. The novel of this model is that it combines the evaluation of method bias with the

evaluation of the single participant performance. Hence, it is possible to address whether it is the method or the participant performance, or both, that needs to be improved.

5.2.3 Analytical quality goals used in EQA for POC INR

Paper II showed that the most common acceptability limit used in EQA schemes for POC INR was a deviation of ±15% from a method specific target value. All EQA organizers used a percent deviation limit from the target value; none used the z-score. However, two organizers reported that they take the uncertainty of the target value into account before calculating the limits; a target *interval* is used instead of a target *value*. This is due to the fact that the target value always is associated with an uncertainty, and to avoid that a larger proportion of results falls outside the limits in low level than in high level.

The acceptability limits used in EQA are often set as a minimum standard, and an EQA result within the limits does not necessarily indicate good analytical quality (107). Consequently, two organizers have divided the acceptability limits into categories (good, acceptable and poor). The reported acceptability limits varied between 12% and 30%, indicating that an EQA result could be assessed as acceptable in one country and unacceptable in another. The limits seem to be narrower in schemes used for educational purposes compared to schemes where participation is mandatory and have legal consequences. Ideally, for similar schemes the acceptability limits should be equal across countries and based on “clinical outcome”. A collective opinion paper (123) state that future effort should be put in harmonizing the EQA acceptability limits across Europe, and that the best approach to determine these limits is by biological variation. It is not known which step in the quality goal hierarchy the reported acceptability limits in paper II are based on. This question could have been included in the questionnaire.

5.2.4 Analytical quality of POC INR methods

Control of the analytical quality of a method does not automatically result in quality improvement (122). IQC and EQA can only monitor and assess the analytical quality. Therefore, it is important that the laboratories take some actions to improve quality if the requirements are not meet. However, it is better to prevent errors than to correct them (44).

Imprecision

The analytical imprecision (repeatability) of Thrombotrack, CoaguChek S and Hemochron Jr. Signature was estimated in paper I, both with lyophilized control materials and with native patient samples. The CV results were 1.9% and 1.5% for Thrombotrack, 7.3% and 5.7% for CoaguChek S, and 9.4% and 8.9% for Hemochron Jr. Signature, respectively. Thrombotrack fulfilled the quality specification based on biological variation for patients on warfarin treatment (CV ≤ 5.4%) (73-76), specification based on clinician's opinion (CV ≤ 4.1%) (79) and specification based on professional recommendations (CV ≤ 5%) (20, 81), whereas CoaguChek S and Hemochron Jr. Signature did not meet these requirements. Other studies found a CV=3.0% for Thrombotrack (124), CV=3.4% (125) and CV=5.5% (126) for CoaguCheck S, and CV=8.5% for Hemochron Jr. Signature (127).

When the analytical precision of a method is poor, it is a challenge to perform IQC (44). One criterion for analytical variation is that it should be less than 1/6 of the maximum allowable error, which is characterized as "six sigma performance" (57). In this situation, it is easy to find a control rule which gives a $P_{ED} > 90\%$ and a $P_{FR} < 5\%$. At four to five sigma performance the control rule(s) must be carefully selected and more effort must be put in the IQC administration. At less than four sigma, the IQC becomes almost impossible and useless because of the low P_{ED} and high P_{FR} (57). This was seen in paper I where CoaguChek S achieved 4.1 sigma performance and Hemochron Jr. Signature achieved 2.6 sigma performance (the calculations are based on the $CV_{within\text{-}day}$ estimates from duplicate patient samples. The maximum

allowable bias was set to 15%, the maximum allowable analytical variation was set to 5%, and the stable bias was ignored). Consequently, no control rule was optimal for these instruments. The corresponding number for Thrombotrack was 15.5 sigma, meaning that all rules were useful. In primary care, it is common that the laboratories use a fixed percent control limit of for example 10%. The control rule $1_{10\%}$ corresponded to 1_{5S} for Thrombotrack, $1_{1.5S}$ for CoaguChek S and 1_{1S} for Hemochron Jr. Signature (calculated based on the $CV_{within\text{-}day}$ estimates from duplicate control samples). Consequently, this control rule was useless for CoaguChek S and Hemochron Jr. Signature with a false error rate of 27% and 29%, respectively (for n=1).

Bias and accuracy

In paper III, the most commonly used POC methods were compared with the most commonly used hospital method in Norway. CoaguChek XS Plus and Simple Simon fulfilled the analytical quality specification of bias less than 6% (based on professional recommendations) (81) and bias less than 2% (based on biological variation for healthy individuals) (72), whereas Thrombotrack did not meet this requirements. However, all three methods fulfilled the ISO (52) requirements of bias less than 0.3 INR in level 2.0-4.5 INR and the inaccuracy requirement of at least 90% of the POC results within 30% of the comparison method. Only the Thrombotrack instrument did not fulfilled the Norwegian inaccuracy requirements of at least 95% of the POC results within 20% of the comparison method (20).

One limitation of the bias estimates in paper III is that there were few samples above 4.5 INR, especially for CoaguChek XS Plus. Bias in high INR level can theoretically be different from bias in low or therapeutic level because the methods are calibrated for the INR level 1.5-4.5 only (13). However, a study which has investigated the analytical quality of CoaguChek XS Plus in level 4.5-8.0 INR found that bias was -0.1% compared with a hospital method (128), which were similar to the biases found in paper III. A review of the analytical quality of POC INR methods for patient self-

testing performed by Christensen and Larsen in 2012 (54) showed that the reported biases for CoaguChek XS Plus varied from -0.40 INR to 0.13 INR. A recent publication found a positive bias of 0.25 INR for CoaguChek XS Plus in the therapeutic level (129). This discrepancy in results could indicate that the comparison methods used in the different studies were not harmonized.

Paper II showed that the CoaguChek instruments were the most commonly used POC INR methods in Europe. The Thrombotrack and Simple Simon instruments are most common in the Nordic countries and only a few studies have investigated the analytical quality of these instruments. The review by Christensen and Larsen did not include Thrombotrack and Simple Simon because these instruments are not suitable for patient self-testing (they are wet chemistry and some laboratory experience is therefore required). However, the Scandinavian evaluation of laboratory equipment for primary health care (SKUP) found that Simple Simon had a bias of +0.13 INR and 98% of the samples were within the quality specification of 20% compared with a hospital method (130). Another study showed that Simple Simon had a bias of +5.1% compared to a hospital method (131). These two studies indicate that the harmonization between Simple Simon and the hospital methods in Norway is acceptable. The bias of Thrombotrack has been reported by SKUP (124) to be +0.64 INR, and only 40% of the results were within the quality specification of 15% compared with a hospital method. These findings were similar to another study (132) which found a bias of +0.54 INR and 57% of the results were within the quality specification of 0.5 INR compared to a hospital method. After the SKUP study in 2000, the manufacturer changed the traceability chain of the thromboplastin used on the Thrombotrack instruments, in order to improve harmonization with the Norwegian hospital methods (see 1.3.4). Even though paper III showed that the bias has been smaller (+0.2 INR) the harmonization is still not satisfactory.

Already in 1959, P. A. Owren claimed that widely differing results could be seen by using different thrombplastins (133). The INR system was therefore developed in order to standardize the analysis (12). The principle is that the ISI value should correct for the sensitivity of different thromboplastins and the INR value should

therefore be independent of the thromboplastin used. It has been shown that different thromboplastins have different sensitivities (134). However, even if the ISI corrects for large differences between thromboplastins and methods, INR discrepancies can still be seen (11). It is important to remember that calculation of INR is based on mean values, and single patient results can therefore deviate much from one method to another. Several publications have reported poor agreement between methods (8, 117, 135), and this indicates that more effort should be put in the harmonization process. Principally, when the methods are not harmonized common reference intervals cannot be used (136). Therefore, different therapeutic intervals for INR monitoring could have been established for different methods.

5.2.5 Recommendations

Some recommendations for analytical quality control of POC methods can be proposed based on the findings in paper I, II and III and other publications. The evidence of the following recommendations is given as references, but level of evidence and strength of recommendation have not been addressed.

Internal quality control

First of all, there is little published evidence that IQC of POC methods actually leads to quality improvement. However, it is recommended that POC users should perform IQC (104, 137, 138), and that the in-built electronic controls are not sufficient (104). IQC of POC INR methods can be performed by using control materials or by using the split sample procedure (89). However, paper I showed that the probability to detect true errors was higher when a control material was used compared to the split sample procedure. This is true as long as the imprecision of the control materials is lower than the imprecision of the split samples. The split sample procedure used as IQC (i.e. when one patient sample is analysed both on a POC method and on a comparison method, and the deviation between these methods is compared against a

control limit) is therefore not recommended because of the various factors that can influence the single split sample result. Split sample may, however, be useful in situations where there are no control materials available or when the control materials are poor, meaning that the imprecision is larger using the control material compared to using patient samples. However, in such cases only large errors can be detected. The split sample procedure should be restricted to method comparison studies, in which the primary care laboratory can estimate the stable analytical bias of their method compared to a designated comparison method (e.g. in cases where the primary care laboratory starts with a new analysis or change instrument/method). When this bias is established, IQC with the use of suitable control materials should be used to detect *changes* in bias.

Recommendation 1*: IQC of POC methods should be performed by using control materials suitable for the methods (104, 137, 138) (paper I) and not by the split sample procedure (paper I).*

The control rules should be chosen based on the prevalence of medically important errors (100). If the prevalence is high, a control rule with high P_{ED} should be preferred, and if the prevalence is low, a control rule with low P_{FR} should be preferred(100). The prevalence of errors was, however, not investigated in this thesis. The control frequency should be related to the potential risk of harm of the patient, and it should be increased when the risk is increased (104). A higher control frequency is required for methods with large analytical imprecision compared to methods with low analytical imprecision. This is because the probability of early error detection is higher when the number of control samples is increased (59), and the number of erroneously patient results reported are thus depending on the time interval between each control measurement (139). This was seen in paper I where the CoaguChek S and Hemochron users would have to perform IQC more frequent than the Thrombotrack users to detect the same size of systematic error.

Recommendation 2*: The IQC frequency should be related to the prevalence of medically important errors and potential risk of harm to the patient (100, 104).*

Recommendation 3*: POC methods with poor analytical quality should be controlled more frequently than POC methods with good analytical quality (paper I).*

Note that the recommendations given from paper I are based on detecting changes in systematic errors. However, there is no reason to believe that the conclusions should be different for detecting an increase in random errors.

External quality assessment

It is recommended that POC users should participate in an EQA scheme whenever available (104, 137, 138). This recommendation is also valid for POC INR methods (89, 108). However, paper II showed that only a few countries in Europe offer an EQA scheme for POC INR methods and different types of schemes are provided. No recommendation can be given regarding which type of EQA scheme for POC INR methods leads to best quality improvements, as this was not examined in this thesis. However, the availability and organization of such schemes were investigated in paper II and the results showed that the most commonly used control material was lyophilized plasma with peer-group target values. Participation in EQA schemes is considered useful even when the control materials are non-commutable and peer-group target values are used. In such schemes, the primary care laboratories can compare their result with others with the same POC method and the between-laboratory variation can be assessed. In addition, participation has an educational aspect with focus on quality improvement. It is assumed that participation in such schemes will improve the quality, but the evidence for this assumption is limited. No quality improvement can be achieved simply by participate in EQA; unacceptable results must be followed by corrective actions.

Recommendation 4*: Users of POC methods should participate in an EQA scheme whenever available (104, 137, 138) (paper II).*

EQA schemes with non-commutable control materials and peer-group target values can be considerably improved by implementing the proposed model presented in

paper III, and it has been recommended that EQA organizers should implement this model (140). As a consequence, the primary care laboratories will get more information about the analytical quality of their POC method. By implementing the proposed model, the schemes become more useful without the need to distribute native samples to all participants. For the EQA organizers the proposed model might be simpler and easier to carry out than traditional method comparison studies because they will often have limited access to patients. In addition, several different methods can be compared to the same designated comparison method, meaning that biases of different methods can be compared and harmonization between methods can be assessed. Lot-to-lot-evaluations can also be performed using this model. The model can in principle be implemented for all constituents, but for practical reasons the sample stability should for example be at least 24 hours (the samples must be transported from the primary care laboratories to the comparison method). This model is thus not suitable for e.g. glucose. The expert primary care laboratories and the designated comparison method should be carefully chosen in order to establish valid bias estimations. The analytical quality specifications should also be carefully selected. By using this model, EQA organizers should be able to advise the participants against using poor methods.

Recommendation 5*: EQA organizations should implement the proposed EQA model presented in paper III in order to improve the utility of EQA schemes that use non-commutable control materials (140).*

6. MAIN CONCLUSIONS

This thesis and its associated papers aimed to evaluate and suggest improvements of the analytical quality control of INR methods used in primary care, and the main conclusions were:

- *Internal quality control*: The probability of detecting systematic errors was higher when the primary care laboratories used lyophilized control materials compared to using patient split samples, for all the investigated control rules. The probability of false alarms was, however, the same. IQC of POC INR methods should therefore be performed by using control materials and not by the split sample procedure.

- *External quality assessment*: Most European countries do not offer EQA schemes for POC INR methods. Only 12 EQA organizations in nine European countries reported that they provide this service. There is a vide variation in how these schemes are organized, but the most common is to use lyophilized control materials, establish peer group target values, use an acceptability limit of 15% and distribute four samples per year.

- *External quality assessment*: A new EQA model was developed in order to improve EQA schemes that use non-commutable control materials. In this model, native patient samples were used in combination with non-commutable (lyophilized) control materials in order to assess both the POC method bias and the single-participant performance. By using this model the primary care laboratories will get more information about the analytical quality of their POC method.

7. FUTURE PERSPECTIVES

Analytical quality control of POC methods faces many challenges. Some of these are addressed in this thesis (exemplified for POC INR methods), but many remain unsolved. An important question is whether or not analytical quality control really leads to quality improvement. Simply performing analytical quality control does not automatically improve the quality. Studies are needed to investigate how the primary care laboratories handle an unacceptable control result and examine if these corrective actions actually leads to quality improvement that will benefit the patient. Such studies are important but may be challenging to conduct. Another important aspect in analytical quality control is the quality of the control materials. No control system is of any value if the control results do not reflect the patient results. A deviating control result should indicate that the patient results will deviate in the same way. However, if the control material is not commutable this assumption may not be valid. Future studies should investigate the commutability of control materials.

The EQA model presented in this thesis could be further developed. The suggested number of expert laboratories and patient samples could be further evaluated and optimized in order to obtain valid bias estimates and at the same time keep the model easy to perform on a regular basis. The EQA organizers could for example collaborate in performing this model so that the workload is reduced. The model could in addition be modified to fit other laboratory methods that are not POC methods. Non-commutable control materials are often used in EQA schemes for hospital laboratories and implementation of this model could improve the evaluation of traceability and inter-method harmonization for several different constituents. Noklus will implement the proposed model in the near future for INR in primary care and maybe also for other constituents.

REFERENCES

1. McMillin GA, Vazquez SR, Pendleton RC. Current challenges in personalizing warfarin therapy. Expert Rev Clin Pharmacol 2011;4:349-62.

2. Lind M, Fahlen M, Kosiborod M, Eliasson B, Oden A. Variability of INR and its relationship with mortality, stroke, bleeding and hospitalisations in patients with atrial fibrillation. Thromb Res 2012;129:32-5.

3. Hirsh J, Dalen JE, Anderson DR, Poller L, Bussey H, Ansell J, et al. Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. Chest 1998;114:445S-69S.

4. Oden A, Fahlen M. Oral anticoagulation and risk of death: a medical record linkage study. BMJ 2002;325:1073-5.

5. White PJ. Patient factors that influence warfarin dose response. J Pharm Pract 2010;23:194-204.

6. Quick AJ. The prothrombin in hemophilia and obstructive jaundice. J Biol Chem 1935;109:73-84.

7. Owren PA. A quantitative one-stage method for the assay of prothrombin. Scand J Clin Lab Invest 1949;1:81-3.

8. Jackson CM, Esnouf MP, Lindahl TL. A critical evaluation of the prothrombin time for monitoring oral anticoagulant therapy. Pathophysiol Haemost Thromb 2003;33:43-51.

9. Horsti J. Comparison of quick and owren prothrombin time with regard to the harmonisation of the International Normalised Ratio (INR) system. Clin Chem Lab Med 2002;40:399-403.

10. Jackson CM, Esnouf MP. Has the time arrived to replace the quick prothrombin time test for monitoring oral anticoagulant therapy? Clin Chem 2005;51:483-5.

11. Poller L. International Normalized Ratios (INR): the first 20 years. J Thromb Haemost 2004;2:849-60.

12. WHO. Expert committee on biological standardization - Requirements for thromboplastins and plasma used to control oral anticoagulant therapy. World Health Organization, Geneva Techical Report Series 889, 48th Report 1999.

13. CLSI. Procedures for Validation of INR and Local calibration of PT/INR Systems; Approved Guideline. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, USA, CLSI document H54-A 2005.

14. van den Besselaar AM, Chantarangkul V, Tripodi A. Thromboplastin standards. Biologicals 2010;38:430-6.

15. van den Besselaar AM, Tripodi A. Is there a need for replacement of the International Reference Preparation for Thromboplastin, Bovine, Combined (OBT/79)? J Thromb Haemost 2005;3:2365-6.

16. van den Besselaar AM, Barrowcliffe TW, Houbouyan-Reveillard LL, Jespersen J, Johnston M, Poller L, Tripodi A. Guidelines on preparation, certification, and use of certified plasmas for ISI calibration and INR determination. J Thromb Haemost 2004;2:1946-53.

17. Poller L, Keown M, Ibrahim S, van den Besselaar AM, Roberts C, Stevenson K, et al. Comparison of local International Sensitivity Index calibration and 'Direct INR' methods in correction of locally reported International Normalized Ratios: an international study. J Thromb Haemost 2007;5:1002-9.

18. Lindahl TL, Egberg N, Hillarp A, Odegaard OR, Edlund B, Svensson J, et al. INR calibration of Owren-type prothrombin time based on the relationship

between PT% and INR utilizing normal plasma samples. Thromb Haemost 2004;91:1223-31.

19. Hillarp A, Egberg N, Nordin G, Stigendal L, Fagerberg I, Lindahl TL. Local INR calibration of the Owren type prothrombin assay greatly improves the intra- and interlaboratory variation. A three-year follow-up from the Swedish national external quality assessment scheme. Thromb Haemost 2004;91:300-7.

20. Trydal T, Bolann B, Brosstad F, Sandberg S, Sandset PM, Stavelin A, et al. [Analysis of prothrombin time in primary health care]. Tidsskr Nor Laegeforen 2005;125:2048-9.

21. ISO/DIS17511. In vitro diagnostics medical devices - Measurement of quantities in samples of biological origin - Metrological traceability of values assigned to calibrators and control materials. CEN - European Committee for Standardization, Brussels, Belgium Draft prEN ISO 17511 2000.

22. Leichsenring I, Plesch W, Unkrig V, Kitchen S, Kitchen DP, Maclean R, et al. Multicentre ISI assignment and calibration of the INR measuring range of a new point-of-care system designed for home monitoring of oral anticoagulation therapy. Thromb Haemost 2007;97:856-61.

23. ITC. Hemochron Jr. Signature Whole Blood Microcoagulation Systems. Package Insert. http://www.itcmed.com/uploads/literature/fb5934_0307.pdf (Accessed 14 May 2013).

24. Roche. CoaguChek S Systems. Package Insert. http://www.apsfa.org/docs/Package%20Insert%20for%20CCk%20and%20CCkS.pdf (Accessed 14 May 2013).

25. Zafena. Simple Simon® PT, Product documentation. http://www.zafena.se/English/SimpleSimon (Accessed 14 May 2013).

26. Axis-Shield. Thrombotest (PT-INR). http://www.axis-shield.com/thrombotest-pt-inr (Accessed 14 May 2013).

27. Lundberg GD. Acting on significant laboratory results. Jama 1981;245:1762-3.

28. Plebani M, Favaloro EJ, Lippi G. Patient safety and quality in laboratory and hemostasis testing: a renewed loop? Semin Thromb Hemost 2012;38:553-8.

29. Lippi G, Guidi GC, Mattiuzzi C, Plebani M. Preanalytical variability: the dark side of the moon in laboratory testing. Clin Chem Lab Med 2006;44:358-65.

30. Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem 2002;48:691-8.

31. O'Kane MJ, McManus P, McGowan N, Lynch PL. Quality error rates in point-of-care testing. Clin Chem 2011;57:1267-71.

32. Favaloro EJ, Lippi G, Adcock DM. Preanalytical and postanalytical variables: the leading causes of diagnostic error in hemostasis? Semin Thromb Hemost 2008;34:612-34.

33. CLSI. Collection, Transport, and Processing of Blood Specimens for Testing Plasma-Based Coagulation Assays and Molecular Hemostasis Assays; Approved Guideline - Fifth Edition. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, USA, CLSI document H21-A5 2008.

34. Brigden ML, Graydon C, McLeod B, Lesperance M. Prothrombin time determination. The lack of need for a discard tube and 24-hour stability. Am J Clin Pathol 1997;108:422-6.

35. Lippi G, Guidi GC. Effect of specimen collection on routine coagulation assays and D-dimer measurement. Clin Chem 2004;50:2150-2.

36. Lippi G, Salvagno GL, Montagnana M, Lima-Oliveira G, Guidi GC, Favaloro EJ. Quality standards for sample collection in coagulation testing. Semin Thromb Hemost 2012;38:565-75.

37. Adcock DM, Kressin DC, Marlar RA. Minimum specimen volume requirements for routine coagulation testing: dependence on citrate concentration. Am J Clin Pathol 1998;109:595-9.

38. Goswami B, Tayal D, Chawla R, Mallika V. Pre-analytical factors that influence the interpretation of prothrombin time in the clinical laboratory: one year experience in a super speciality hospital in India. Clin Chim Acta 2009;410:93-4.

39. van Geest-Daalderop JH, Mulder AB, Boonman-de Winter LJ, Hoekstra MM, van den Besselaar AM. Preanalytical variables and off-site blood collection: influences on the results of the prothrombin time/international normalized ratio test and implications for monitoring of oral anticoagulant therapy. Clin Chem 2005;51:561-8.

40. Favaloro EJ. Stability of coagulation assays performed in plasma from citrated whole blood transported at ambient temperature: only a part of the story. Thromb Haemost 2008;99:1122-3.

41. Eilertsen H, Dagre AT, Ulveland K. [Peroral anticoagulant treatment--are the INR values reliable?]. Tidsskr Nor Laegeforen 2000;120:618.

42. Baglin T, Luddington R. Reliability of delayed INR determination: implications for decentralized anticoagulant care with off-site blood sampling. Br J Haematol 1997;96:431-4.

43. Blann AD, Fitzmaurice DA, Lip GY. Anticoagulation in hospitals and general practice. BMJ 2003;326:153-6.

44. Petersen PH, Ricos C, Stockl D, Libeer JC, Baadenhuijsen H, Fraser C, Thienpont L. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. Eur J Clin Chem Clin Biochem 1996;34:983-99.

45. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accred Qual Assur 2007;12:45-7.

46. VIM. International Vocabulary of Metrology - Basic and General Concepts and Associated Terms, 3th edition. JCGM 200:2008 (E/F) 2008.

47. Theodorsson E. Validation and verification of measurement methods in clinical chemistry. Bioanalysis 2012;4:305-20.

48. Barwick VJ, Prichard EE. Eurachem Guide: Terminology in Analytical Measurement - Introduction to VIM 3. First Edition. www.eurachem.org (Accessed 21 June 2013).

49. Jacobson AK. Warfarin monitoring: point-of-care INR testing limitations and interpretation of the prothrombin time. J Thromb Thrombolysis 2008;25:10-1.

50. Poller L, van den Besselaar AM, Jespersen J, Tripodi A, Houghton D. The European Concerted Action on Anticoagulation (ECAA): field studies of coagulometer effects on the ISI of ECAA thromboplastins. Thromb Haemost 1998;80:615-23.

51. Poller L, Keown M, Chauhan N, van den Besselaar AM, Tripodi A, Shiach C, Jespersen J. European Concerted Action on Anticoagulation. A multicentre calibration study of WHO international reference preparations for thromboplastin, rabbit (RBT/90) and human (rTF/95). J Clin Pathol 2005;58:667-9.

52. ISO17593. Clinical laboratory testing and *in vitro* medical devices - Requirements for *in vitro* monitoring systems for self-testing of oral anticoaglant therapy. International Organization for Standardization, Geneva, Switzerland 2007;First edition. April 15.

53. JCTLM. Joint Committee for Traceability in Laboratory Medicine. http://www.bipm.org/jctlm/ (Accessed 15 May 2013).

54. Christensen TD, Larsen TB. Precision and accuracy of point-of-care testing coagulometers used for self-testing and self-management of oral anticoagulation therapy. J Thromb Haemost 2012;10:251-60.

55. Linnet K, Boyd J. Selection and analytical evaluation of methods - with statistical techniques. In: Tietz Fundamentals of Clinical Chemistry (6th ed), Burtis CA, Ashwood ER, Bruns (red), Elsevier Saunders 2008:201-28.

56. Fraser CG. General strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999;59:487-90.

57. Klee GG, Westgard J. Quality Management. In: Burtis CA, Ashwood ER, Bruns DE, editors Tietz Fundamentals of Clinical Chemistry, 6th ed Elsevier Saunders, Oxford, UK 2008:249-62.

58. Westgard JO. Error budgets for quality management--practical tools for planning and assuring the analytical quality of laboratory testing processes. Clin Lab Manage Rev 1996;10:377-81, 84-93, 96-403.

59. Westgard JO, Groth T. Power functions for statistical control rules. Clin Chem 1979;25:863-9.

60. Bolann BJ, Asberg A. Analytical quality goals derived from the total deviation from patients' homeostatic set points, with a margin for analytical errors. Scand J Clin Lab Invest 2004;64:443-50.

61. Klee GG. Establishment of outcome-related analytic performance goals. Clin Chem 2010;56:714-22.

62. ISO13528. Statistical methods for use in proficency testing by interlaboratory comparisons. International Organization for Standardization, Geneva, Switzerland First edition 2005.

63. Cuskun A, Unsal I, Serteser M, Fraterman A. Towards standardization of external quality assessment schemes by using bias values based on biological variation. Accred Qual Assur 2009;14:547-52.

64. Shermock KM, Connor JT, Lavallee DC, Streiff MB. Clinical decision-making as the basis for assessing agreement between measures of the International Normalized Ratio. J Thromb Haemost 2009;7:87-93.

65. Hemkens LG, Hilden KM, Hartschen S, Kaiser T, Didjurgeit U, Hansen R, et al. A randomized trial comparing INR monitoring devices in patients with anticoagulation self-management: evaluation of a novel error-grid approach. J Thromb Thrombolysis 2008;26:22-30.

66. Petersen JR, Vonmarensdorf HM, Weiss HL, Elghetany MT. Use of error grid analysis to evaluate acceptability of a point of care prothrombin time meter. Clin Chim Acta 2010;411:131-4.

67. Fraser CG, Petersen PH. Analytical performance characteristics should be judged against objective quality specifications. Clin Chem 1999;45:321-3.

68. Dot D, Miro J, Fuentes-Arderiu X. Within-subject and between-subject biological variation of prothrombin time and activated partial thromboplastin time. Ann Clin Biochem 1992;29 (Pt 4):422-5.

69. Wada Y, Kurihara M, Toyofuku M, Kawamura M, Iida H, Kayamori Y, et al. Analytical goals for coagulation tests based on biological variation. Clin Chem Lab Med 2004;42:79-83.

70. Rudez G, Meijer P, Spronk HM, Leebeek FW, ten Cate H, Kluft C, de Maat MP. Biological variation in inflammatory and hemostatic markers. J Thromb Haemost 2009;7:1247-55.

71. Costongs GM, Bas BM, Janson PC, Hermans J, Brombacher PJ, van Wersch JW. Short-term and long-term intra-individual variations and critical

differences of coagulation parameters. J Clin Chem Clin Biochem 1985;23:405-10.

72. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491-500.

73. van den Besselaar AM, Fogar P, Pengo V, Palareti G, Braham S, Moia M, Tripodi A. Biological variation of INR in stable patients on long-term anticoagulation with warfarin. Thromb Res 2012;130:535-7.

74. Lassen JF, Brandslund I, Antonsen S. International normalized ratio for prothrombin times in patients taking oral anticoagulants: critical difference and probability of significant change in consecutive measurements. Clin Chem 1995;41:444-7.

75. van Geest-Daalderop JH, Pequeriaux NC, van den Besselaar AM. Variability of INR in patients on stable long-term treatment with phenprocoumon and acenocoumarol and implications for analytical quality requirements. Thromb Haemost 2009;102:588-92.

76. Van Geest-Daalderop JH, Kraaijenhagen RJ, Van Der Meer FJ, Van Den Besselaar AM. Intraindividual variation of the International Normalized Ratio in patients monitored with a recombinant human thromboplastin. J Thromb Haemost 2010;8:1641-2.

77. Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. Clin Chem 2011;57:1334-6.

78. Gowans EM, Hyltoft Petersen P, Blaabjerg O, Horder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. Scand J Clin Lab Invest 1988;48:757-64.

79. Kristoffersen AH, Thue G, Sandberg S. Postanalytical external quality assessment of warfarin monitoring in primary healthcare. Clin Chem 2006;52:1871-8.

80. Kristoffersen AH, Thue G, Ajzner E, Claes N, Horvath AR, Leonetti R, et al. Interpretation and management of INR results: a case history based survey in 13 countries. Thromb Res 2012;130:309-15.

81. Kvalitetssikring og kvalitetskrav til laboratoriemedicinske aktiviteter i almen praksis. Regionernes Lønnings- og Takstnævn (RLTN) og Praktiserende Lægers Organisation (PLO), Konsensusdokument, Danmark 2010.

82. Fitzmaurice DA, Gardiner C, Kitchen S, Mackie I, Murray ET, Machin SJ. An evidence-based review and guidelines for patient self-testing and management of oral anticoagulation. Br J Haematol 2005;131:156-65.

83. CLSI. Evaluation of Matrix Effects; Approved Guideline - Second Edition. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, USA, CLSI document EP14-A2 2005.

84. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta 2003;327:25-37.

85. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev 2007;28:139-47.

86. Brion E, Lessinger JM, Gould N, Leyendecker J, Ferard G. Evaluation of commutability of control materials. Clin Chem Lab Med 2002;40:625-30.

87. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples--five year experience from the Netherlands. Clin Chim Acta 2012;414:234-40.

88. Baadenhuijsen H, Kuypers A, Weykamp C, Cobbaert C, Jansen R. External Quality Assessment in The Netherlands: time to introduce commutable survey specimens. Lessons from the Dutch "Calibration 2000" project. Clin Chem Lab Med 2005;43:304-7.

89. NCCLS. Point-of-Care Monitoring of Anticoagulation Therapy; Approved Guideline. The National Committee for Clinical Laboratory Standards, Wayne, Pennsylvania, USA, NCCLS document H49-A 2004.

90. van den Besselaar AM, Hoekstra MM. Long-term stability and reproducibility of CoaguChek test strips. Thromb Haemost 2005;93:1189-92.

91. van den Besselaar AM, Hoekstra MM, van der Meer FJ. Quality control of CoaguChek test strips in the Netherlands. Thromb Haemost 2007;97:323-4.

92. Kitchen S, Kitchen DP, Jennings I, Woods TA, Walker ID, Preston FE. Point-of-care International Normalised Ratios: UK NEQAS experience demonstrates necessity for proficiency testing of three different monitors. Thromb Haemost 2006;96:590-6.

93. Solvik UO, Stavelin A, Christensen NG, Sandberg S. External quality assessment of prothrombin time: the split-sample model compared with external quality assessment with commercial control material. Scand J Clin Lab Invest 2006;66:337-49.

94. Kessler P. External quality assessment of prothrombin time testing on POCT monitor. [Czech]. Klinicka Biochemie a Metabolismus 2010;18 (4):210-3.

95. RCPA. Royal College of Pathologists of Australasia (RCPA) Quality Assurance Program (QAP). http://www.rcpaqap.com/ (Accessed 15 May 2013).

96. CAP. College of American Pathologists. http://www.cap.org (Accessed 15 May 2013).

97. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem 1977;23:1857-67.

98. Bhattacharyya GK, Johnson RA. Statistical Concept and Methods. John Wiley & Sons, Inc (New York, Chichester, Brisbane, Toronto, Singapore) 1977;Chapter 6:165-86.

99. Westgard JO. Selecting appropriate quality-control rules. Clin Chem 1994;40:499-501.

100. Westgard JO, Hyltoft Petersen P, Groth T. The quality-costs of an analytical process: 2. A test yield formulation of the predictive value quality-costs model. Scand J Clin Lab Invest 1984;172:229-36.

101. Westgard JO, Klee GG. Quality Management. In: Burtis CA, Ashwood ER, Bruns DE, editors Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, 4th ed Elsevier Saunders, Oxford, UK 2006:485-529.

102. Neubauer AS. The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation. Clin Chem 1997;43:594-601.

103. Winkel P. Statistisk kvalitetsudvikling i klink og laboratorium. Ingeniøren bøger, Copenhagen 2002:149-63.

104. Cooper G, DeJonge N, Ehrmeyer S, Yundt-Pacheco J, Jansen R, Ricos C, Plebani M. Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality. Clin Chem Lab Med 2011;49:793-802.

105. Kitchen DP, Kitchen S, Jennings I, Woods T, Walker I. Quality assessment in point-of-care coagulation testing. Semin Thromb Hemost 2008;34:647-53.

106. ISO15189. Medical laboratories - particular requirements for quality and competence. International Organization for Standardization, Geneva, Switzerland 2007.

107. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670-80.

108. Perry DJ, Fitzmaurice DA, Kitchen S, Mackie IJ, Mallett S. Point-of-care testing in haemostasis. Br J Haematol 2010;150:501-14.

109. Poller L, Keown M, Chauhan N, van den Besselaar AM, Tripodi A, Shiach C, Jespersen J. European concerted action on anticoagulation. Quality assessment of the CoaguChek Mini and TAS PT-NC point-of-care whole-blood prothrombin time monitors. Clin Chem 2004;50:537-44.

110. Kitchen S, Kitchen D, Jennings I, Woods T, Walker I. Quality assessment of CoaguChek point-of-care international normalized ratio monitors: a note of caution. Clin Chem 2007;53:1555-6; author reply 6-7.

111. Murray ET, Jennings I, Kitchen D, Kitchen S, Fitzmaurice DA. Quality assurance for oral anticoagulation self management: a cluster randomized trial. J Thromb Haemost 2008;6:464-9.

112. Tripodi A, Bressi C, Carpenedo M, Chantarangkul V, Clerici M, Mannucci PM. Quality assurance program for whole blood prothrombin time-international normalized ratio point-of-care monitors used for patient self-testing to control oral anticoagulation. Thromb Res 2004;113:35-40.

113. Blis CI. Statistics in Biology - Statistical methods for research in the natural sciences. Volume one. McGraw-Hill Book Company (New York, St Louise, San Francisco, Toronto, London, Sydney) 1967;Chapter 10:231-79.

114. Burnett RW. Accurate estimation of standard deviations for quantitative methods used in clinical chemistry. Clin Chem 1975;21:1935-8.

115. UN. United Nations Statistics Division. http://unstats.un.org (Accessed 10 May 2011).

116. Spannagl M, Dick A, Reinauer H. External quality assessment schemes in coagulation in Germany: between regulatory bodies and patient outcome. Semin Thromb Hemost 2007;33:259-64.

117. Horsti J, Uppa H, Vilpo JA. Poor agreement among prothrombin time international normalized ratio methods: comparison of seven commercial reagents. Clin Chem 2005;51:553-60.

118. CLSI. Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline - Second Edition (Interim Revision). Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, USA, CLSI document EP09-A2-IR 2010.

119. Kitchen D, Murray E, Kitchen S, Jennings I, Fitzmaurice D, Woods T, Walker I. External quality assessment for prothrombin time/international normalised ratio using point-of-care devices: split-sample or conventional external quality assessment using control samples? Scand J Clin Lab Invest 2007;67:673-5; author reply 6-7.

120. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. Clin Chem 2011;57:76-83.

121. Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine chemistry analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. Arch Pathol Lab Med 1998;122:587-608.

122. Libeer JC, Baadenhuijsen H, Fraser CG, Petersen PH, Ricos C, Stockl D, Thienpont L. Characterization and classification of external quality assessment

schemes (EQA) according to objectives such as evaluation of method and participant bias and standard deviation. Eur J Clin Chem Clin Biochem 1996;34:665-78.

123. Adams O, Cooper G, Fraser C, Hubmann M, Jones G, Plebani M, et al. Collective opinion paper on findings of the 2011 convocation of experts on laboratory quality. Clin Chem Lab Med 2012;50:1547-58.

124. SKUP. Report from an evaluation of Thrombotrack/ Thrombotest organised by SKUP. Report no. SKUP/2000/8*. http://www.skup.nu (Accessed 11 April 2013).

125. Christensen TD, Larsen TB, Jensen C, Maegaard M, Sorensen B. International normalised ratio (INR) measured on the CoaguChek S and XS compared with the laboratory for determination of precision and accuracy. Thromb Haemost 2009;101:563-9.

126. Attermann J, Andersen NT, Korsgaard H, Maegaard M, Hasenkam JM. Precision of INR measured with a patient operated whole blood coagulometer. Thromb Res 2003;110:65-8.

127. Maddox JM, Bogo PH, McGregor E, Pippard MJ, Kerr R. Quality assurance for point-of-care testing of oral anticoagulation: a large-scale evaluation of the Hemochron Junior Signature Microcoagulation System. Int J Lab Hematol 2009;31:142-50.

128. Lawrie AS, Hills J, Longair I, Green L, Gardiner C, Machin SJ, Cohen H. The clinical significance of differences between point-of-care and laboratory INR methods in over-anticoagulated patients. Thromb Res 2012;130:110-4.

129. Hur M, Kim H, Park CM, La Gioia A, Choi SG, Choi JH, et al. Comparison of International Normalized Ratio Measurement between CoaguChek XS Plus and STA-R Coagulation Analyzers. Biomed Res Int 2013;2013:213109.

130. SKUP. Report from an evaluation of Simple Simon organised by SKUP. Report no. SKUP/2007/57*. http://www.skup.nu (Accessed 11 April 2013).

131. Solvik UO, Petersen PH, Monsen G, Stavelin AV, Sandberg S. Discrepancies in international normalized ratio results between instruments: a model to split the variation into subcomponents. Clin Chem 2010;56:1618-26.

132. Ignjatovic V, Barnes C, Newall F, Hamilton S, Burgess J, Monagle P. Point of care monitoring of oral anticoagulant therapy in children: comparison of CoaguChek Plus and Thrombotest methods with venous international normalised ratio. Thromb Haemost 2004;92:734-7.

133. Owren PA. Thrombotest. A new method for controlling anticoagulant therapy. Lancet 1959;2:754-8.

134. Testa S, Morstabilini G, Fattorini A, Galli L, Denti N, D'Angelo A. Discrepant sensitivity of thromboplastin reagents to clotting factor levels explored by the prothrombin time in patients on stable oral anticoagulant treatment: impact on the international normalized ratio system. Haematologica 2002;87:1265-73.

135. Tripodi A, Chantarangkul V, Legnani C, Fronton R, Testa S. Discrepancy of the international normalized ratio observed in the external quality assessment survey: a cause for concern. J Thromb Haemost 2012;10:714-6.

136. Petersen PH, Rustad P. Prerequisites for establishing common reference intervals. Scand J Clin Lab Invest 2004;64:285-92.

137. ISO22870. Point-of-care testing (POCT) - Requirements for quality and competence. International Organization for Standardization, Geneva, Switzerland 2005;Final Draft.

138. CLSI. Quality Management: Approaches to Reducing Errors at the Point of Care; Approved Guideline. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, USA, CLSI document POCT07-A 2010.

139. Parvin CA. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. Clin Chem 2008;54:2049-54.

140. Horowitz GL. Proficiency testing matters. Clin Chem 2013;59:335-7.