Acknowledgments

These studies have been financially supported by grants from the Danish Heart Foundation (nos. 97-1-5-80a-22454 and 97-2-6-76A-22522) and from the Faculty of Health Sciences, University of Aarhus. Roche Diagnostics provided the test carriers and the thromboplastin preparations used in the studies.
“There is Nothing so Practical as a Good Theory”

Kurt Lewin
German-born American social psychologist
1890-1947
Preface to first version

This Ph.D. thesis represents the first part of a research project, which aims at establishing a reliable method for routine monitoring of oral anticoagulant therapy.

On 1 February 1996, Niels Trolle Andersen, associate professor at the Department of Biostatistics, University of Aarhus, introduced me to J. Michael Hasenkam, professor in experimental heart surgery at Skejby Sygehus and the Institute of Experimental Clinical Research, University of Aarhus. At this meeting, Michael introduced me to a very exciting problem: the task of establishing an optimal monitoring system for patients on self-managing oral anticoagulant therapy at the newly established center at Skejby Sygehus.

I was very happy with being given the opportunity to use mathematical and statistical methods to solve a highly relevant and meaningful medical challenge, and I was even more delighted, when Niels and Michael later offered themselves as supervisors for a Ph.D. project!

In the course of 1996 and in the beginning of 1997 I worked on the preparation of the Ph.D. project. In this important phase, Michael introduced me to the scientific circles of anticoagulant therapy, and he was of immense help in the demarcation of the goals of the project. Also Michael Væth, professor at the Department of Biostatistics, University of Aarhus, was most obliging and helpful with discussing the project formulation. On many occasions I have exploited his extensive knowledge and formidable intelligence.

During the whole project I have worked very closely with Niels, whose never failing support and sharp intuition has been a safety net, preventing me from being hurt on occasions when the going got tough. Also, the completion of this project would have been difficult without the continuous support and critical comments from Michael Hasenkam.

I thank Jørgen Ingerslev, MD, DMSc, Centre for Haemophilia and Thrombosis, Skejby Sygehus for introducing me to the theory of anticoagulant therapy, Hanne Grønnesby, RN, Department of Cardiothoracic and Vascular Surgery, Skejby Sygehus, Marianne Maegaard, RN, The Anticoagulation Center at Skejby Sygehus, Jan Møller, chemist, Anne-Marie Holm Peder-
sen and Marianne Lyngbak, laboratory technicians, Department of Clinical Biochemistry, Skejby Sygehus for their patience with my inquiries and silly questions, Jim Hill, principal scientist and Winfried Plesch, Ph.D., Roche Diagnostics, Indianapolis, USA, and Mannheim, Germany for valuable discussions and smooth cooperation, and Karin Kynde, pharmacist, Department of Clinical Chemistry, Roskilde County Hospital and coordinator of the Danish quality assurance program for INR for help with understanding the INR system.

I thank my friend and colleague Søren Lundbye-Christensen, associate professor at the Department of Mathematics, Aalborg University for lots of fun and for introducing me to the theory of state-space models. I also thank my other colleagues in Jutland’s Society for Longitudinal Studies and in our study group on contemporary inference in state-space models for fruitful meetings.

Throughout the whole project I have been offered excellent working conditions at the Department of Biostatistics. I wish to thank my friends and colleagues at the department for always being ready for an inspiring discussion of everything ranging from the possibility of statistical inference of causality to the quality of different brands of roofing felt.

I thank my friends Kim Vestberg, MD., and Carsten Wiuf, Ph.D., for their continuous interest in and willingness to discuss the philosophical and scientific questions of my life.

This work would not have been possible without the loving care, patience, and support from my wife, Karin, and my sons, Jesper and Martin. Thank you for giving me the freedom to occasionally be a nerd and to work at the most peculiar times of the day.

Finally, I extend a heartfelt thank to mum and dad, Grete and Ernst, and to my sister and brother, Else and Lars, for their wholehearted and unconditional support throughout my whole life.

Jørn Attermann
November 1999
Preface to second version

This second version of the dissertation has been prepared under close consideration of the suggestions given by the evaluation committee after submission of the first version.

Specifically, the following changes have been made:

- The introductions specific for the three substudies have been included in the general introduction or the background chapter.

- A chapter with background information on measurement methods and standards for measuring coagulant activity has been included. Several authors are cited in extenso, since this provides the best presentation of their ideas and observations.

- A much larger set of references has been included.

- The detailed discussions, including references, specific for the three substudies, have been included in a general discussion chapter.

- A general conclusion chapter has been included.

- It has been further emphasized that the structure of the dissertation is that of a monograph, i.e. “a learned treatise on a small area of learning”. The monograph is based on three substudies and each of these has moreover been published or submitted for publication in an international peer-reviewed periodical. The layout of the monograph has been changed in order to give a more “smooth” appearance.

The core text remains, however, essentially unchanged.

This dissertation is the result of an interdisciplinary work with elements from three major areas with very different traditions: epistemology, natural science, and medical science. Moreover, the medical parts of it include elements from at least four different medical specialties: cardiothoracic surgery, cardiology, clinical chemistry and hematology.
Since 1984, when I was matriculated as a medical student, I have taken an interest in all the three major areas, but my main occupation in the past 12 years has been with the quantitative epistemological discipline of statistics, which deals with the collection, analysis, interpretation, and presentation of numerical data. The formal foundation for theoretical statistics is mathematical statistics, in which I took a master’s degree (MSc) in 1995.

Concurrently, in the past nine years, and especially since 1995, when I was first employed at the Department of Biostatistics, I have been occupied with biostatistics, which focuses on the development and use of statistical methods to solve problems and to answer questions that arise in human biology and medicine. My experience has been within various scientific disciplines including different medical specialties, and in every case I have found that knowledge of the subject matter of the particular area is an important prerequisite in order to be able to communicate with the specialist.

A remarkable feature of many of the scientific problems I have encountered as a biostatistician is their similarity! It is striking, and fascinating, how different scientific areas have encountered the same epistemological challenges and how they independently have invented almost the same (good or bad) strategies for dealing with the analysis and interpretation of data. Another striking feature is the difference in language and traditions, which hinders the different areas in communicating their epistemological problems and discoveries to each other. With this in mind I consider it the principal goal of biostatistics to 1) collect epistemological and cognitive problems and solutions from different subject matter areas of research; 2) distill the problems and solutions through the language of mathematical statistics; 3) present, analyze and discuss the problems and solutions in a general setting within the mathematical-statistical scientific community; and 4) translate possible solutions to the problems to the language and traditions of the particular subject matter disciplines.

A major challenge as a biostatistician is the translation of solutions and suggestions from very general terms to terms which are recognized by the subject matter specialist. I hope that this dissertation will be judged on the scientific usefulness of its applications of general statistical methodology, rather than on its exact compliance with the language and traditions of the subject matter disciplines, which it involves.

Jørn Attermann
May 2000
## Contents

1 Introduction ................................................. 1

2 Background .................................................. 5
   2.1 Plasma based methods ................................. 5
       2.1.1 Early results .................................. 5
       2.1.2 Towards a standard ............................. 11
       2.1.3 Studies on analytical variation ................. 17
       2.1.4 The International Normalized Ratio ............ 20
       2.1.5 Danish quality controlled laboratories .......... 36
   2.2 Capillary whole blood coagulometers ................. 37

3 Materials and methods of empirical studies .............. 43
   3.1 Precision of home coagulometer ..................... 43
       3.1.1 Study subjects ................................ 43
       3.1.2 Patient self-testing of INR by the home coagulometer . 43
       3.1.3 Laboratory measurements of INR ................ 43
       3.1.4 Study design .................................. 44
       3.1.5 Statistical methods ............................ 45
   3.2 Validity of the INR standard ......................... 45
       3.2.1 Plasma samples ................................ 45
       3.2.2 Measurement methods ........................... 46
       3.2.3 Study design .................................. 47
       3.2.4 Statistical methods ............................ 47

4 Results of empirical studies ............................. 53
   4.1 Precision of home coagulometer ..................... 53
       4.1.1 Precision ...................................... 53
       4.1.2 Comparison of home coagulometers with routine hospital laboratories ......................... 55
   4.2 Validity of INR standard ............................. 56
       4.2.1 Marginal results ............................... 56
4.2.2 The linearity hypotheses ........................................ 57
4.2.3 INR values ......................................................... 59

5 Accuracy of INR estimates: mathematical derivation 65
  5.1 A motivating example: The calibration of the International
      Reference Preparation RBT/79 .................................. 65
  5.2 A contemporary model for INR ................................... 67
      5.2.1 A semi-deterministic model ............................... 67
      5.2.2 Extension to a statistical model ......................... 68
  5.3 Distribution of the INR estimator ............................... 70
      5.3.1 A simulation study ....................................... 77
  5.4 The interlaboratory variation ................................... 82
  5.5 Example: Estimates derived from RBT/90 ....................... 87
      5.5.1 \( t_1 \) versus \( t_0 \): Calibration of RBT/79 against IRP 67/40 87
      5.5.2 \( t_2 \) versus \( t_1 \): Calibration of RBT/90 against RBT/79 . 91

6 Discussion 93
  6.1 Precision of home coagulometers ............................... 93
  6.2 Validity of the INR standard ................................... 96
  6.3 Accuracy of INR estimates ..................................... 100

7 Conclusion and perspectives 103

A Some additional statistical results 105
  A.1 Plots of differences between measurement methods ........... 105
  A.2 The lognormal distribution .................................. 106

B SAS programs 107
  B.1 Plotting and testing in calibrations ......................... 107
      B.1.1 The CALIB macro ...................................... 108
      B.1.2 The EREG macro ...................................... 113
  B.2 Estimating interlaboratory variance ......................... 117
  B.3 Simulation of log(log(INR)) ................................. 122

Summary 125

References 129

Abbreviations 143
1

Introduction

In the beginning of the 20th century it was known by veterinarians and cattle raisers that cattle fairly frequently and for no apparent reason bleed considerably and perhaps fatally. In 1921 F.W. Schofield and L.M. Roderick determined independently that this hemorrhagic disease arises from eating improperly cured hay or silage made from the common sweet clovers. The hemorrhagic agent, dicoumarol\textsuperscript{1}, was isolated and synthesized in the late thirties and early fourties by Karl P. Link and associates from the Wisconsin Alumni Research Foundation. The effect of dicoumarol given orally to humans was reviewed and studied in 1942 by Edgar W. Allen et al. [1]. They concluded that “clinical and experimental studies strongly suggest the value of administration of dicoumarin in preventing intravascular thrombosis”. They also noted that hemorrhage may occur when dicoumarol has greatly prolonged the clotting time and recommended that “dicoumarin should be administered only when its effect can be determined by repeated calculations of the prothrombin time”.

Today, life-long oral anticoagulant therapy (OAT) is offered to patients with increased risk of thrombosis, e.g. patients with artificial heart valves, atrial fibrillation or deep venous thrombosis. It is estimated that in 1992 in the Nordic countries (Sweden, Norway, Finland, Iceland, and Denmark) 0.3 – 0.5\% of the population was undergoing daily anticoagulant therapy [2]. The measurement of coagulant activity is indispensable to the management of OAT, and with the introduction of the INR standard in 1983, a common platform was established for reporting test results. Moreover, with the development of small portable whole blood coagulometers it has become possible to offer self-management of OAT to selected patients.

\textsuperscript{1}Allen et al. use the name dicoumarin for the compound 3,3’-methylene-bis-(4-hydroxycoumarin). Other names for the same compound are dicumarin, dicumarol and dicoumarol.
A necessary prerequisite for evaluating the quality of any particular concept of OAT is an investigation of the accuracy of INR estimates obtained by the measurement method belonging to that concept. The International Organization for Standardization (ISO) [3] defines accuracy as “the closeness of agreement between a test result and the accepted reference value”; trueness as “the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value”; and precision as “the closeness of agreement between independent test results obtained under stipulated conditions.” Thus, in order to evaluate the accuracy, both trueness and precision must be investigated.

In order to evaluate the accuracy of INR estimates obtained by patients on self-managing OAT, we started out by studying the precision of patients’ own measurements of INR performed at home by using CoaguChek portable whole blood coagulometers (substudy 1). This is the first published study on the precision of routine INR measurements performed at home by patients on self-managing OAT.

It turned out to be more difficult to evaluate the trueness of INR estimates. Previous reports suggested that the accuracy of routine laboratory INR estimates is not sufficient for them to act as accepted reference values [4, 5]. A necessary condition for accuracy of a given method is that it fulfills the fundamental linearity assumptions of the INR standard. Since the Department of Clinical Biochemistry at Skejby Sygehus used the Nycotest method for routine estimation of INR, and the CoaguChek coagulometers were calibrated by the Hepato Quick method by the manufacturer, we chose to examine the validity of the INR standard for these two methods. If the assumptions of the INR standard were found to be valid for both methods, this would indicate that accurate measurements can be obtained by both CoaguChek and Nycotest, provided that the manufacturer has achieved linearity in the calibration of CoaguChek against Hepato Quick.

The accuracy of INR estimates was further investigated in the third sub-
study. Numerous reports on inconsistent or discrepant INR values [6, 7, 8, 9, 10, 11, 12] suggest that the distributional properties of the INR system are not fully understood. The uncertainty of ISI estimates has previously been discussed by Tomenson (see page 30), who loosely estimated that for an INR estimated at 3.5, the 95% confidence interval may be as large as 2.47 - 4.96 for routine measurements. The accuracy of INR estimates obtained directly by using IRPs with the recommended method of WHO has, however, never been precisely estimated. We propose a realistic statistical model for INR estimates, based on close examination of reports on calibration exercises for IRPs [13, 14, 15, 16, 17, 18, 19], as well as on our self-experienced calibration exercise in the second sub-study. The proposed statistical model is based on
the contemporary reality, i.e. how things actually are being done. Thus, the estimates of INR, ISI etc., which are being examined, are not necessarily the optimal ones, but the ones, which are used in practice. The purpose of this substudy was to derive a closed and accurate description of the distribution of INR estimates given the true INR in order to 1) identify the components of variance, and 2) provide the basis for a cost-benefit analysis for variance reduction. This description may be used to calculate the variance of the INR estimates derived from the various international reference preparations, and this may offer an explanation of the discrepant and “inexplicable” INR values, which are frequently reported in the scientific literature.

The specific hypotheses to be investigated in the substudies were:

**Substudy 1** The precision of patient’s own measurements of INR performed at home on CoaguChek coagulometers is sufficient to allow for self-management of OAT.

**Substudy 2** For any pair of methods (Nycotest, Hepato Quick, and CRM 149S) the following holds true:

- \( H_1 \): The relationship between the true logarithms of PTs of abnormal plasmas is linear.
- \( H_2 \): The point corresponding to the means of the normal log PTs lies on the line of the abnormal plasmas.

Secondarily, there is no clinically significant deviation between INR measured by the three methods.

**Substudy 3** Based on a realistic statistical model, the statistical distribution of INR estimates, given a true INR value, can be adequately described to allow for good estimates of accuracy and precision.

First, in chapter 2, the necessary background information for the three substudies is presented. In chapter 3 the material and methods for the two empirical studies (the first and second substudy) is presented. The results of these two studies are presented in chapter 4. Chapter 5 contains the mathematical derivation of the statistical distribution of INR estimates. The results of the empirical studies and the derived formulae for the accuracy of INR estimates are discussed in chapter 6. Finally, a conclusion and further perspectives are presented in chapter 7.
2

Background

2.1 Plasma based methods

2.1.1 Early results

The one-stage method of Quick et al.

In 1931 L.M. Roderick and A.L. Schalk found that the delayed coagulability of the blood in cattle fed with spoiled clover involved a reduction in the prothrombin content of the blood and that this diminution paralleled the delay in the coagulation time [1].

Later, in 1935, Armand J. Quick, Margaret Stanley-Brown and Frederic W. Bancroft [20] investigated four substances, which according to a theory attributed to J. Bordet (1920), P. Morawitz (1914) and E. Wöhlsch (1929) were the essential components in blood clotting. According to this theory clotting proceeds in two phases:

1. prothrombin + thromboplastin + calcium $\rightarrow$ thrombin

2. fibrinogen + thrombin $\rightarrow$ fibrin

Thus, only four constituents are essential for clotting: prothrombin, thromboplastin, calcium, and fibrinogen. From a review of the literature and from own experiments Quick et al. concluded that “fibrinogen and calcium were probably rarely responsible for disturbances in blood clotting and therefore required little consideration other than a brief summary of their relation to clotting”. Thus, “attention must be centered on prothrombin and thromboplastin”.

The rationale behind the determination of the prothrombin time (PT) is beautifully described by Quick et al.: “From the equation:
prothrombin + thromboplastin + calcium = thrombin

it can be postulated that if both calcium and thromboplastin be kept constant, the rate of thrombin formation becomes proportional to the concentration of prothrombin. Incidentally, Eagle [1934-1935] has demonstrated that the amount of thrombin formed is independent of calcium and platelets but varies directly with the plasma factor (prothrombin). By employing oxalated plasma and recalcifying with a fixed and optimal quantity of calcium, this factor is made constant. On adding an active preparation of thromboplastin, a maximum acceleration in coagulation time is obtained even with an exceedingly small amount, and, significantly, further addition will not shorten the clotting time . . . . Thus, with the addition of a fixed quantity of calcium and an excess of thromboplastin, prothrombin is the only variable and the clotting time of oxalated plasma under these conditions can be considered a direct measure of the prothrombin concentration of the blood."

Quick et al. proceeded with describing in details a method for prothrombin determination: “Nine cubic centimeters of blood, withdrawn rapidly and with special precaution to avoid trauma, are promptly and thoroughly mixed with 1 cc. of M/10 sodium oxalate, and centrifuged at a low speed for 5 minutes. Of this plasma 0.1 cc. is transferred to a dry clean test tube (13 by 100 mm.), and mixed with 0.1 cc. of thromboplastin solution. Then, 0.1 cc. of M/40 calcium chlorid is added, and the tube quickly shaken, and placed in a water bath kept at 37C. The exact time required for the formation of a solid clot is recorded.”

The preparation of thromboplastin required special attention: “For carrying out the new prothrombin test successfully, it is necessary to prepare an active and stable thromboplastin. Rabbit brain was found to be the most satisfactory source, since it is not only very active, but it can be preserved in the dry state for weeks with little deterioration in activity. The preparation is as follows: the brain of a freshly killed rabbit is freed of the larger superficial blood vessels, washed, then ground to a paste, and spread in a thin layer on a plate glass, or a flat dish. After thoroughly drying at 37C., the material is removed from the plate, and preserved in a stoppered container. By mixing 0.2 gm. of this material with 0.3 cc. of 0.85% sodium chlorid, and incubating for 15 minutes, an emulsion is obtained which has practically maximum activity, i.e., it will when added to human plasma cause clotting in 22 to 25 seconds. Any preparation which fails to show this activity is discarded. A fresh emulsion prepared from rabbit brain, thymus or lung, and heated to 60C. for 15 minutes to inactivate any prothrombin which may be present, likewise brings about clotting in 22 seconds, thus suggesting that the thromboplastin in these 3 tissues is identical. For the test either the
fresh or the dry preparation may be employed. Attempts to purify or isolate the active constituent invariably led to a diminution of activity, and since the product contains no impurity which would interfere with the present problem no further attempts at purification were deemed necessary.”

This method was later slightly modified, apparently by T.B. Magath in 1939, in the following way: At the same instant of adding calcium chloride a stop clock is started and the tube is carefully agitated and then tilted at intervals to determine the moment at which a semi-solid clot is formed [21]. This procedure is known as the manual tilt tube technique.

In their paper from 1942, Allen et al. [1] stated: “Dicoumarin is a hemorrhagic agent which may produce dangerous hemorrhage unless its administration is controlled rigidly. The amount of drug administered must depend on the prothrombin time. Hence, at present, dicoumarin should not be given unless the prothrombin time can be determined. It is our opinion that the Quick method, as modified by Magath, is satisfactory.”

**Prothrombin percentage**

Very importantly, Quick et al. [20] noted that with thromboplastin prepared from a series of more than 20 rabbits from New York City, the clotting time was 22 to 25 seconds, whereas thromboplastin prepared from rabbits obtained in Milwaukee had an “equally consistent but distinctly great activity”, since the clotting time obtained with this thromboplastin was 16 to 17 seconds. Thus, different thromboplastins gave different clotting times, even though they were prepared from the same species and by the same methods.

In 1945 Hurn et al. [21] performed a dilution experiment where PT was determined in various dilutions of normal plasma with different thromboplastins. The resulting dilution curves (one for each thromboplastin) could then be used for obtaining the Prothrombin Percentage (PP) corresponding to a given PT obtained by a given thromboplastin. The authors emphasized that the Prothrombin Percentage is a measure of activity and not necessarily reflects the actual percentage of normal prothrombin concentration. Many years later, in 1967, Biggs & Denson, [22] criticized the Prothrombin Percentage for being inaccurate: “Most of the commercial methods provide a calibration curve which converts the results to “percentage of normality”. These dilution curves do not ensure any uniformity between methods of assessing the degree of abnormality, because the sensitivity of the methods to the coumarin defect is not truly represented by these dilution curves. The curves are made by testing the clotting-times by the one-stage technique of dilutions of normal plasma, and they vary very considerably according to the diluent used, to the nature of the thromboplastin, and to other details of
2. BACKGROUND

technique. To many the dilution curve has seemed such an arbitrary scale of measurement that the use of simple clotting-time ratios have been preferred. . . . The ratio method has the advantage that it does not carry any unjustified implication of scientific validity.” In 1975 Beeser et al. [23] reported from a German proficiency testing program that the variation between laboratories and thromboplastins was much larger when using percentage activity than seconds or PT ratio.

Owren’s method

Following the discovery of factor V in 1943 the world became more complicated. In 1949 Owren wrote [24] in a critique of Quick’s one-stage method and the two-stage method: “In the various one-stage technics, oxalated (or citrated) plasma (or whole blood) is mixed with thromboplastin and calcium, and the clotting time determined. . . . Quick has insisted that the clotting time under these conditions depends entirely on the amount of prothrombin present and consequently may be used as a measure of prothrombin concentration.

The discovery of the fifth clotting factor or factor V (Owren, 1943) offered a complication in this method. It was shown that the conversion of prothrombin to thrombin needs the presence of factor V, and that the rate of this conversion depends on the amount of factor V present. The ‘prothrombin time’ of Quick is not therefore a measure of prothrombin concentration but varies with the absolute as well as relative concentrations of both prothrombin and factor V in the plasma. Calculations of percentage of prothrombin time using this method are erroneous. Decreased amount of factor V also gives a prolonged Quick’s ‘prothrombin time’. Furthermore, Quick’s method is insensitive. Significant prolongation of the clotting time by this method does not occur until the concentration of factor V has dropped below 50% of the normal. Increased concentrations of prothrombin or factor V cannot be detected. . . .

In the two-stage method [of Warner, Brinkhous & Smith], thromboplastin and calcium are added to dilutions of defibrinated oxalated (or citrated) plasma, and the maximum amount of thrombin formed is determined by the clotting activity on fibrinogen. When factor V is reduced, the conversion of prothrombin to thrombin proceeds slowly. During the time needed for the maximal thrombin titer to occur, a part of the thrombin form is inactivated by the antithrombin, and consequently the thrombin amount recorded will be too small. By absence of factor V no prothrombin is detected by this method.”

As a solution to this, Owren suggested to measure the PT by using throm-
boplasrin with added bovine plasma, which is freed from prothrombin. In such a plasma, the concentration of factor V is high, and thus factor V deficiency will not disturb the estimation of prothrombin time. The commercial thromboplasrin preparation, Thrombotest, is based on this idea.

**Normotest**

The Normotest method was devised by Owren & Strandli in 1969 [25] to “assay the true activity of three clotting factors: II, VII and X, independently of the presence of endogenous coagulation inhibitors, in order to serve as a liver function test.”

Although it was not designed to measure PT in patients on OAT, the method is mentioned here, since it actually has been used for that kind of measurements. “The insensitivity of Normotest to inhibitors has been achieved by selecting a thromboplasrin preparation of inherent low activity which has been processed and freed from serum clotting factors and intermediates by the same procedure as used for Thrombotest. Although serum contaminants may compensate for inhibitors (and thereby increase the insensitivity) they have to be completely removed, because they will produce a false acceleration of the clotting reaction in the absence of inhibitors in the tested plasma and secondly, their presence will seriously hamper the standardization of the reagent and the reproducibility of results from one batch to another. . . .

The Normotest reagent is a lyophilized preparation containing optimal amounts of a standardized inhibitor-insensitive thromboplasrin, factor V, fibrinogen and calcium.”

**An early informal calibration procedure**

As noted by Quick and many others, the PT depends to a large degree on the preparation of the thromboplasrin used. In 1945 Margaret Hurn et al. [21] wrote: “The factor having the greatest influence on the activity of thromboplasrin prepared from rabbit brain is whether or not the brain has been extracted with acetone. . . . Unfortunately thromboplastins prepared according to the same general method do not always possess similar activity. This is true regardless of the time which has elapsed between the preparation of the dried brain and preparation of the thromboplasrin. It would not be profitable to enumerate every possible reason why individual preparations of thromboplasrin may vary in activity. It should be realized, however, that the fineness of the powdered brain, the length of time and the temperature at which the saline extract is incubated, the amount of mixing prior to, during,
and after incubation, and the period of centrifugation each may have an effect on the potency of the final thromboplastin. ... Regardless of precautions taken, one may occasionally obtain a thromboplastin so different in activity from other preparations used that it is advisable to discard it.”

Hurn et al. demonstrated how the PT increases with increasing concentration of rabbit brain mixture in the thromboplastin. They noted that “there seems to be little change from day to day in thromboplastins prepared as outlined” in their paper. In order to obtain consistency between prothrombin times measured by different thromboplastin preparations, Hurn et al. recommended a sort of informal calibration procedure, where each fresh preparation of thromboplastin is checked with a previously standardized thromboplastin.

“Thromboplastin was considered to be ‘standardized’ if it gave a normal prothrombin time of from seventeen to nineteen seconds, if it gave prothrombin times that were comparable (not more than 5 per cent variation) to those obtained with thromboplastins which had been previously used, and if, in addition, it would give a prothrombin time of approximately thirty-five seconds when tested with a 20 per cent dilution of a selected normal plasma. ... Sometimes it was necessary to reduce the amount of dried brain weighed out in order to obtain a thromboplastin of the desired activity.”

Hurn et al. also noted a linear relationship between prothrombin times in seconds for different thromboplastins and thus the possibility of obtaining an expected value for thromboplastin A by measuring the prothrombin time using thromboplastin B. Their calculations did not, however, correct for measurement errors.

Calibration by PT ratio

In 1966 Rosemary Biggs and K. W. E. Denson [26] reported a calibration exercise using different types of thromboplastin preparations. In an earlier study they found an approximate linear relationship between the prothrombin time ratios (PTR) between patient and normal plasma using two different types of thromboplastin. Due to this linear relationship they found it reasonable to define the sensitivity (denoted Thromboplastin Sensitivity Ratio, TSR, in Bangham, Biggs, et al. 1973 [27]) of a given thromboplastin preparation as the PTR between patient and normal plasma that would be obtained by that thromboplastin, given that the PTR was 2.0 using Thrombotest.

In a preliminary study it was found that the variability due to reconstitution depended on whether the acetone-dried brain extract was homogenized or not. For brain extract, which was gently stirred, the sensitivity was robust to dilution. For homogenized brain extract, however, the sensitivity decreased markedly when the brain extract was diluted (fig. 2 in [26]). Thus
the authors “thought it likely that better results would be obtained using ampouled preparations in which there would be little scope for variations in the technique of reconstitution.”

A series of inter-laboratory comparisons of PT ratios were performed using 7 different types of thromboplastin preparations. Twenty eight centers from 11 countries using 10 different anticoagulant drugs participated in the trial.

“Each centre was asked to test 5 samples from anticoagulant treated patients together with a normal sample on 4 different days. Each centre thus compared 20 anticoagulant samples with 4 different normal samples.”

The authors found no systematic differences attributable either to the type of drug used or the geographical location.

“For each centre the results for the different methods, expressed as clotting time ratios, were plotted to give a correlation with the thrombotest results. The results of all 4 days observations were included together.” Pairwise plots of PTRs obtained by Thrombotest against the other thromboplastins give the impression of linear relationships. Moreover, the variance tends to increase with increasing magnitude of the ratios.

The authors noted that the reproducibility error increases with decreasing sensitivity, and thus thromboplastins with low sensitivity should not be used. Thromboplastins with high sensitivity may, however, become very inconvenient with a high degree of anticoagulation because of the long clotting times. The authors pointed out that “the ideal would appear from our observations to be a method comparable in sensitivity to Thrombotest but with the shortest possible clotting time with normal plasma. If the ox brain of the Thrombotest reagent is replaced by rabbit brain much shorter clotting times are obtained with normal plasma and the sensitivity approaches that of Thrombotest. Preliminary experiments with a reagent of this sort suggest that it might prove generally very satisfactory.”

Contrary to expectations the authors found no difference in variability of sensitivity due to ampouled or non-ampouled thromboplastin preparations.

2.1.2 Towards a standard

Biggs & Denson [26] concluded: “The work has now reached a stage at which advance must depend on a number of practical decisions.

1. A reference standard thromboplastin preparation is required. There are a number of decisions which should be reached about this. a) What sort of preparation should be made? There is in our view much to be said for preferring the type of preparation described by Denson. Does
2. BACKGROUND

the committee [for Haemostasis and Thrombosis] support this view? b) By whom should the reference standard be made? c) To whom should the standard be available?

2. The committee should accept a definition of a good method for carrying out the one-stage prothrombin time and the definition should include the requirement that, when tested with a number of samples from anticoagulant treated cases, the ratio of abnormal to normal clotting times should not be less than that obtained using the standard preparation by a factor to be determined when the standard thromboplastin is available. This factor would be so arranged that a ratio of 1.6 in comparison with a Thrombotest ratio of 2 times should be a minimum accepted sensitivity.

3. A mechanism whereby commercial preparations can be tested and approved should be set up. . . .

4. Immediately a reference preparation becomes available, commercial firms should be urged to standardise their preparations in terms of the international reference preparation. . . . ”

Commenting on this, A.S. Littell [28] lined up four desirable properties of an assay system:

1. “The dose-response curve should be a straight line over a wide range of the dose scale.

2. The variation of responses for the same dose should be small, but should be the same for different doses.

3. The slope should be large, or more specifically, the ratio of the standard deviation to the slope should be small.

4. The above three should be constant over time and in different laboratories.”

As noted by Littell, linearity and variance homogeneity can frequently be obtained within a given system by appropriate transformation.

Littell continued: “Comparing the features of the design of this trial with the properties desired in a bioassay, there was just one material that was tested at two doses [i.e. dilutions] (the normal), so any deviation from linearity cannot be detected. Although duplicate clotting time measurements were made at most centers, these were not independently repeated prothrombin tests, so no estimate of the standard deviation of the response can be made.
2.1. PLASMA BASED METHODS

The slope, without an estimate of its error and without the standard deviation of the response, is of very limited comparative value. Thus, according to the usual bioassay models, there was nothing that could be interpreted from these trials regarding the relative precision of the 7 assays based on the 7 thromboplastin preparations or the reproducibility from center to center. If this study had been designed a little differently - say using on each day two normals at each of 3 dilution levels with independent replicates - there would have been comparative data on all 4 of the properties a bioassay system should have.”

In principle Littell is right in proposing the four desirable properties, except that the assumptions of linearity and homogeneity of variances can be relaxed with modern methods. It seems, however, that Littell does not recognize that in this case the clinicians are not interested in the dose-response relationship between the dose (the degree of coagulate activity, which happens to be unobservable) and the response (the clotting time), which is otherwise typical in bioassays. What Littell thinks of as the dose appears to be the degree of dilution of normal plasma, which does not truly reflect the coumarin defect, as noted by Biggs & Denson. Being a bit wise after the event, it seems that what the clinicians want to know in the present case is the clotting time (possibly the PT ratio or another transformation) that would have been obtained for a given plasma sample, had it been analyzed using a certain common reference thromboplastin under certain further specified reproducibility conditions.

In 1967 Biggs & Denson proposed a standard preparation of thromboplastin containing absorbed ox plasma and using human brain as a source of thromboplastin [22]. “The principle of the method of standardization depends quite simply on the comparison of the method in question with a reference preparation of known sensitivity. The two methods are used on a number of samples from patients receiving anticoagulant therapy and the results compared as in” a table of measurements of approximately 20 plasma samples from patients receiving anticoagulant therapy. When furthermore a dilution curve was prepared for the standard thromboplastin, it would be possible approximately to convert PTRs obtained by the test thromboplastin to Percentage Prothrombin as would have been obtained by the standard thromboplastin. The authors noted that “the method could very well be used to standardize commercial thromboplastins, which would then be issued with an approved scale.” The authors made an important point in noting that “a preparation and method which is insensitive to the defect caused by coumarin drugs can never give satisfactory results. At present the majority of preparations are calibrated more by their reaction to saline dilutions of normal plasma than by their sensitivity to the defect caused by
coumarin drugs. Thus some commercial preparations are very insensitive. When the first reference preparation is available it would be an advance for a minimum acceptable sensitivity to be established by standards institutions. Such a concept would encourage the preparation of satisfactory commercial preparations. ... A single scale properly applied at different centres would ensure safety and uniform dosage for a patient moving from one place to another and would greatly improve the standard of clinical trials carried out at more than one centre."

Problems with standardization

The problems with standardization were very well summarized by Owren & Strandli in their 1969 paper [25]: “Different thromboplastin time methods record different clotting activities. Some methods are sensitive to inhibitors, others not. Some thromboplastins contain a high and others a low contamination with serum factors and coagulation intermediates which compensate, wholly or partly, for the lack of factors VII and X in the tested plasma. The sensitivity to specific factors varies with the species specificity of the thromboplastin. It is not surprising, therefore, that the elaborate work in recent years devoted to the search for a reliable technique of comparing results with different methods has so far been largely unsuccessful. Uniform results of comparison can not be obtained as long as different methods measure different things. This can be exemplified by comparing results of the testing of dicumarol plasma with Thrombotest (which is sensitive to inhibitor) and Normotest (which is insensitive to inhibitor). The difference in results by testing of the same plasma varies from one patient to another, because of individual variations in the amount of inhibitor.”

Owren [29]: “The method we originally used for factor X determination was based on an artificially prepared factor X deficient ox plasma, ox brain thromboplastin and a final dilution of the test plasma in the reaction mixture of 1:5. This method has later been found to be highly sensitive to the PIVKA-inhibitor¹ and it records the resultant or residual factor X activity similar to Thrombotest. Loeliger et al. by using human brain thromboplastin, congenitally deficient or artificially prepared factor X deficient plasma and a final dilution of the test plasma of 1:50, found no significant difference in the depression of the activity of factor X as compared with that of factor II and VII in patients on long term anticoagulant therapy. The discrepancy between our results is explained by the fact that methods using human brain thromboplastin and high dilution of the test plasma are highly insensitive to

¹Protein Induced by Vitamin K Absence or antagonists
In the control of anticoagulant therapy it is highly desirable to use methods which are sensitive to inhibitors and reflect the total change of coagulability.

Normotest, like Thrombotest, the P-P method and Quick’s method is a thromboplastin time test. The sensitivity of all these tests to the presence of inhibitors depends on two main factors:

1. The nature and characteristics of the thromboplastin.

2. The dilution of the test plasma in the final reaction mixture.”

In 1974 Fischer & Falkensammer [30] compared Hepato Quick with Normotest and Thrombotest, and found apparently a linear relation between percentage activities. In 1979 Loeliger & van Halem-Visser [31] in a study of 14 types of thromboplastin found that “all of these thromboplastins displayed sufficient factor VII sensitivity, and all were sensitive to activation products. . . . Rabbit brain thromboplastins were the most sensitive preparations and human as well as rabbit diluted the least sensitive.” In 1981 Frost, Lau & Jones [32] performed a comparative study of “2 thromboplastins, the Australasian Reference Thromboplastin, and Simplastin, in a large group of patients on long-term Warfarin therapy. 373 individual samples were obtained. Calibration constants were obtained for those patients with prothrombin ratios within the therapeutic range, and for those well outside the therapeutic range, and found to be different. Study of the relationship between the 2 thromboplastins indicates that comparability is linear only within a specified limited range of prothrombin ratios. At the two extreme ends the relationship is curved, suggesting a logarithmic relationship. Attention is drawn to the need of caution in interpretation of corrected ratios calculated on a linear relationship especially when the ratio is above 4.0 as this may have clinical implications.”

**Calibration of five thromboplastins**

In 1973 Bangham et al. [27] reported a trial carried out with five different thromboplastins (69/223, 67/40, 68/434, 70/115, and 70/178) using fresh and freeze-dried plasma. Three laboratories in three countries took part in the study. Fresh plasmas were collected in the participating laboratories on the day the tests were carried out.

“Tests on each plasma were performed in replicate, a mean time calculated and the result expressed as a clotting time ratio, that is the clotting time of the PIVKA-inhibitor. . . .
the tested plasma over the mean clotting time of all normal plasmas tested in the same experiment.

The clotting time ratio with different thromboplastins were plotted against each other according to Biggs & Denson. The best straight line was fitted by eye and the ratio of the test thromboplastin to the reference one read from the graph, for each experiment individually. All the comparisons between different thromboplastins were expressed as thromboplastin sensitivity ratio (TSR), defined as the ratio of test thromboplastin corresponding to the ratio of 2.0 with 67/40 taken as reference.” I.e., 67/40 was by definition assigned a TSR of 2.0.

“A number of experiments from all laboratories had to be discarded because of minor discrepancies in following the instruction procedures. This emphasised once more the great care required in following the procedural instructions; small discrepancies in volumes used and reconstitution procedures may make great differences in calibration results.

One of the participating laboratories also commented on the differences in end-point readings depending upon whether a tilt-tube technique or a Kolle hook were used. These differences became noticeable with abnormal plasmas only, where protracted clot formation occurred. It also appeared that in this laboratory hand methods gave more accurate results for 68/434, whereas machine methods gave more accurate results with thromboplastins 70/115 and 70/178.”

Unfortunately the authors did not perform a thorough statistical analysis; they merely stated the mean and range of TSRs obtained in the different experiments, and it is impossible to evaluate the precision or the accuracy from these numbers. Moreover the estimation of the TSRs were done by eye, and this may introduce a new source of variation. The authors, however, postulated:

“The results of this study confirmed that thromboplastins sensitivity to the defect introduced by coumarins and/or indanediones, can be quantitated in terms of TSR using one thromboplastin as reference. . . . All materials showed excellent stability on accelerated degradation; their stability after reconstitution requires further investigation.

A statistical method for calculating precision (in the form of conventional limits of error) of the comparison of two thromboplastins and to assess the heterogeneity of the results is under study.

The results of this study showed that it is possible to obtain similar TSRs with fresh and freeze-dried patient’s plasma for all thromboplastins except 70/115. When artificially prepared deficient plasmas were used, only TSR for thromboplastin 69/223 in terms of 67/40 was comparable to the TSR obtained with fresh plasma. The TSR for thromboplastin 68/434 was lower
than with patient’s plasma, whereas for thromboplastins 70/115 and 70/178 the TSR was higher than calculated from results using fresh plasma. There are probably several reasons for these differences: thromboplastin 68/434 is known to be sensitive to the inhibitor of prothrombin conversion... present in patients on long term oral anticoagulants. As this inhibitor is absent from the artificially prepared deficient plasma, it can be expected that thromboplastin 68/434 would be estimated to have a lower TSR than when using patient’s plasma.

The results of this study suggest that the plasma carefully collected from patients stabilized on long term anticoagulants and freeze-dried could be used as substrate for calibration of at least some thromboplastin preparations (four of the five thromboplastins investigated in the present study). However the stability of factors II, VII, and X in non-buffered freeze-dried pooled human plasma is not sufficient... for them to be suitable to serve as long-term reference materials.

The artificially prepared abnormal plasmas included in this study were found to be of little use in calibrating the thromboplastin preparations.

2.1.3 Studies on analytical variation

Bowyer et al. (1971) [33] performed a nice study of variation in PT with the following sources: within-run, among-run (day) over 10 non-consecutive days, 2 fibrometers, 4 thromboplastins, and two lyophilized pools of citrated plasma with normal and prolonged PT. Sources of variation was estimated by analysis of variance. Although the analysis apparently was performed on raw PTs, which typically exhibit heteroscedasticity due to increasing variance with increasing levels of measurement, the study clearly demonstrated that there can be significant differences between PTs obtained with different thromboplastins and different types of instrumentation. Only means and standard deviations are given in [33], but from tables 1, 2 and 3 in the paper, the within-run and between-day coefficient of variations (CVs) can be estimated to be of the magnitude 2–3% and 3–4%, respectively.

Leck et al., in 1974 [34], published results from 275 UK hospitals from 4 quality control trials. “Our most recent results confirm the previous finding that there is always considerable variation in the results reported when different hospitals try to measure the prothrombin time ratio for aliquots of the same plasma with the standard thromboplastin. But although there is a considerable tendency for hospitals to obtain consistently high or low ratios for plasmas examined on the same day, this consistency is barely apparent when results collected months apart are compared. We have therefore to abandon the idea that applying a correction based on previous performance
to each hospital’s results might improve their accuracy.”

Koepke et al., in 1975, [35] found that some of the variation was due to incorrect procedures, since many laboratory technicians did not follow the manufacturer’s guidelines. Moreover, the authors discussed the correct concentration of citrate, the effect of hematocrit, and the stability of blood specimens during incubation in room temperature.

**Normalization of Percentage Prothrombin by logarithmic transformation**

Kahan & Norén, in 1975 [36], calculated the coagulation activity, expressed as percentage of normal and as clotting time ratio, in 220 specimens from patients on OAT by three different thromboplastins (Thrombotest, Simplastin-A, and Normotest). They found that the “distribution of percentage values showed a fairly pronounced deviation from normality. After logarithmic transformation, the distribution was normalized, the regression lines between Thrombotest and other tests became parallel, and that between PIVKA-insensitive systems was shifted to a 45 degree line. Logarithmic transformation also stabilized the residual variance. These features make percentage values accessible for treatment according to the standard methods of bioassay statistics.

Attempts to normalize the distribution of ratio values by various transformations were unsuccessful. Formal analysis of data revealed a variation in the proportionality of ratio values with the level of estimated coagulation defect. This may restrict the usefulness of the ratio approach. Logarithmic transformation partly reduced the discrepancy.”

**Standardization recommendations by the WHO in 1977**

In 1977 the World Health Organization (WHO) recommended a standardization procedure [37] based on the observation that plots of PT ratios against each others for different thromboplastins apparently yielded straight lines. WHO recommended the use of three international reference preparations (IRPs) with “calibration constants” (slopes of lines) relative to the IRP 67/40 (human, combined) with a calibration constant defined to be 1:

- Thromboplastin, bovine, combined (coded 68/434): calibration constant = 1.0
- Thromboplastin, human, brain (coded 69/223): calibration constant = 0.90
2.1. PLASMA BASED METHODS

- Thromboplastin, rabbit, brain (coded 70/183): calibration constant = 0.60

The WHO report did not mention how the “calibration constants” should be determined, and random variation was not at all taken into account.

Effect of instrumentation and thromboplastin

In 1981 Evatt et al. [38] analyzed measurements of PT in three different plasma samples by 2580 laboratories. Analysis of variance (ANOVA) was used to show that instrument, as well as thromboplastin, had a significant effect upon observed PT. The authors wrote: “It is unfortunate that the effects of the instruments, the second important laboratory variable affecting the prothrombin time test, have not been studied as thoroughly as have the effects of the thromboplastins.” One of the authors, Donna Brogan, performed a 2-way ANOVA (fixed effects model) with the factors thromboplastin (8 levels) and instrument/method (7 levels). Only 2 of the 56 cells were empty. The ANOVA was performed separately for each of three plasma samples with mean PTs 11.8 s, 15.4 s, and 239 s. For each plasma the interaction between thromboplastins and methods was significant, but the magnitude of the F statistic was small compared with the magnitude of the F statistic for the main effects. Thus, she considered the interaction as not practically or clinically significant and was led to the simpler additive model.

The total CV of the three plasmas were found to be 8%, 11%, and 16%, respectively. Brogan found a funny linear relationship between the main effect of thromboplastins and instruments, respectively, and the mean PT of the plasma sample in question. The main effects estimates stem, however, from a fixed effects model, and the effects are assumed to be constant over time.

In a new study in 1984 by van den Besselaar et al. [39] both instruments and thromboplastins had a highly significant effect on PT. “Relationships based on lyophilized pooled patient plasmas were different from those of artificially depleted plasmas from either manufacturer.”

A year later, Han et al. [40] examined the effects of instrumentation and thromboplastin on PT by ANOVA on logarithmic PTs. They concluded that “for proficiency testing, no advantage was found in the use of either a common thromboplastin or freeze-dried, coumadinized patient plasmas rather than artificially depleted commercial plasmas, except for special purposes.”

A study by Jean M. Thomson et al. in 1983 [41] demonstrated a significant activation of the extrinsic system in blood samples collected by Vacutainer tubes. In contrast, blood collected by a syringe method (plastic or polypropylene syringes with disposable needles) did not show these changes.”
2.1.4 The International Normalized Ratio

Kirkwood's analysis

In 1983 the statistician T.B.L. Kirkwood presented a very thorough analysis of the calibration problem [42]. Kirkwood examined the method recommended by WHO in 1977, which was “based on the empirical observation that if the PT ratios for a set of patients’ plasmas obtained using two different thromboplastins are plotted against one another, an approximately straight line relationship is observed. ... On the reasonable assumption that absence of prolongation of the PT [of normal plasma] with one thromboplastin will correspond with absence of prolongation with the other, the straight line relationship is made to pass through [the origin (1,1)]. The only variable, therefore, is the slope, $b$, of the line. Once $b$ is known, or estimated from the graph, a PT ratio for one thromboplastin may readily be converted into the equivalent ratio for the other using the equation

$$y - 1 = b(x - 1) \quad (1)$$

where $x$ and $y$ refer to PT ratio values on the horizontal and vertical axes, respectively.

The original WHO standardization scheme was based simply on equation (1). The slope $b$ was termed the calibration constant, and when the thromboplastin represented on the horizontal axis was the IRP, $b$ was designated the International Calibration Constant (ICC) of the thromboplastin represented on the vertical axis. The ICC of the primary IRP (67/40) was defined as 1, since this is necessarily the value that would be obtained if the IRP were calibrated against itself (barring experimental error). Thus, any thromboplastin could be compared with the IRP and assigned an ICC, and thereafter any PT ratio could be converted, using equation 1, into the equivalent ratio that would have been obtained if the IRP itself had been used to measure it. To carry out this conversion, equation (1) was rewritten:

$$x = (y - 1)/b + 1. \quad (2)$$

To apply equation (2), the value of the ICC was to be substituted in place of $b$, and the measured PT ratio was to be substituted in place of $y$. The resulting value of $x$ was then the estimate of the PT ratio that would have been obtained if the IRP had been used, and this was termed the International Calibrated Ratio (ICR). By converting all PT ratios into ICRs it was possible, in theory, to regard them as if they were all measured using a single thromboplastin, the IRP. This represented obvious advantages for standardizing the clinical management of oral anticoagulation.
It was recognized that, in practice, the stock of any IRP must be finite and would be exhausted rapidly if it were used to calibrate every thromboplastin directly. This risk was avoided by recommending the calibration of secondary reference preparations, which could then be used in place of the IRP to calibrate other, working preparations. If the ICC of such a secondary reference preparation was ICC$_s$, and the calibration constant of a particular working preparation in terms of the secondary reference preparation was $b_{ws}$, the ICC of the working preparation was to be calculated as

$$\text{ICC}_w = b_{ws} \times \text{ICC}_s.$$ 

It may readily be seen that, apart from extra statistical error arising from the additional calibration step, ICC$_w$ would have been the same as if the working preparation had been calibrated directly against the IRP.

Since the assumption of a straight line relationship through the origin of a PT ratio plot is an empirical one, it should occasion no surprise that exceptions to this pattern have been found [e.g. Kahan & Norén [36]]. These exceptions have arisen generally in cases where thromboplastins of very different sensitivities have been compared and where the calibration constant has been markedly different from 1. Since it is likely that whatever is the true relationship between the two sets of PT ratios it will pass through (1,1), it may be that in such cases this relationship is curved. However, the original WHO scheme cannot meaningfully accommodate either a curvilinear relationship, or a straight line which does not go through (1,1), so any attempt to assign a calibration constant in such cases must be abandoned.

The statistical difficulties with the original WHO scheme stem principally from the fact that the calibration line must be fitted to ratios for which both numerator (patient’s PT) and denominator (normal PT) are subject to error. While the error in the patients’ PTs features straightforwardly in [plots of PT ratios], the error in the normal PTs does not. The latter error, in fact, perturbs the location of the true origin of the calibration relationship which in consequence is not exactly the point (1,1) but some point differing from it by an unknown random deviance. While the effect of this perturbation on the calibration constant is likely to be small, it cannot be ignored and may be especially significant for the assessment of statistical precision.

In view of these difficulties an alternative method of thromboplastin calibration has been proposed which retains the main principles of the Biggs-Denson method, but which will be easier to implement. The basic modification which has been made to the WHO calibration method is to plot, instead of PT ratios, the PTs themselves on logarithmic scales. This alternative was previously considered by Murphy & Denson and M. Hills.
revised plot, the PTs for normal plasmas appear explicitly on the graph as individual data points. This has a number of distinct advantages.

Firstly, the graph shows immediately the number and scatter of the normal data points, which are treated in exactly the same way as those for the patients. The relationship between the PTs for the two thromboplastins is thus shown quite clearly in its primary form. Secondly, it appears from empirical observation that this relationship is either linear or very close to linear, even in cases where the Biggs-Denson model breaks down. Thirdly, since the data points ... are all independent single observations, the statistical operations of fitting the calibration relationship and of estimating its precision are relatively straightforward.

If the relationship between two thromboplastins in the log PT plot is observed to be linear, a calibration scheme directly analogous to the original WHO scheme can be implemented as follows.

The equation of the calibration line is of the form

$$\log(PT_v) = c \log(PT_h) + d$$

where $c$ is the slope, $d$ is the intercept, and $v$ and $h$ signify the vertical and horizontal axes, respectively. ... It may readily be converted into an equation for PT ratio, which has the very simple form

$$y = x^c \quad (3)$$

where $y$ and $x$ are PT ratios for the thromboplastins represented on the vertical and horizontal axes, respectively. ... In the original WHO scheme, the ICC was defined as the slope, $b$, when the IRP was represented on the horizontal axis. Equation (1) then gave the PT ratio ($y$) for the other thromboplastin in terms of the PT ratio ($x$) for the IRP. In practice, however, equation (1) was to be used the other way around, that is, $x$ was to be calculated from $y$ (equation (2)). ... It would be more straightforward if the calibration relationship were inverted by representing the IRP on the vertical axis. ...

To avoid confusion of the revised scheme with the original one new terms have also been defined in place of ICC and ICR. The new terms are:

**International Sensitivity Index (ISI)** to denote the slope $c$ of the calibration line in a log PT plot where the IRP (67/40) is represented on the vertical axis.

**International Normalized Ratio (INR)** to denote the PT ratio $y$ calculated from equation (3) when $c$ is the appropriate ISI.”
2.1. PLASMA BASED METHODS

Kirkwood emphasized that ordinary least squares (OLS) is inappropriate for the estimation of the slope of the calibration line of the log PTs, owing to the error in both the $x$- and the $y$-variables. Estimation by OLS leads to underestimation of the slope and this may lead to a serious bias, especially in serial calibrations. Kirkwood stated that orthogonal regression can be used. However, “implicit in the orthogonal regression method is an assumption that the statistical error on both axes is similar in magnitude, and this cannot presently be tested (since it includes both biological and experimental variation, and the former cannot be statistically replicated). However, similarity of the two errors seems reasonable a priori and the result is only marginally affected if the assumption is not strictly fulfilled. In the unlikely event that subsequent results reveal gross invalidity of this assumption, the method can readily be modified.”

It is important to realize that the INR here is defined as the PT ratio that would have been obtained, had the reference thromboplastin IRP 67/40 been used and the calibration scheme and the methods of measurement must reflect this. Moreover, the calibration relationship should always represent tests as they are actually carried out in practice. Kirkwood also considered how the calibration scheme should be implemented:

“There are three distinct levels at which the operation of the calibration and standardisation scheme needs to be considered: the hospital or clinic, the manufacturer, and the standards or control authority. In organisational terms, the scheme is a hierarchical one, starting from the primary reference thromboplastin, the IRP, and fanning out through secondary reference preparations and through manufacturers’ production lots to the individual PT ratio determination on patients.”

Each working thromboplastin preparation, used by the laboratory for individual PT ratio determinations, “needs to be assigned an ISI by calibration against a reference preparation of known ISI, whether it be manufactured for commercial sale or produced non-commercially for “in-house” use. Calibration of a thromboplastin is, in general, more precise when comparisons are made between similar preparations of the same species. For this reason, the WHO has established IRPs for bovine and rabbit thromboplastins calibrated in terms of the human IRP (67/40). . . . Wherever possible, calibration should be done only between thromboplastins of the same species.” Kirkwood provided, however, no evidence for this view.

“The ISI for the working thromboplastin preparation is obtained by multiplying the ISI for the reference preparation by the calibration slope. The statistical imprecision of [the assigned ISI] is composed not only of statistical variation in the calibration slope, but also of any error in the calibration of the reference preparation.
For the reason just noted, the calibration of a secondary reference thromboplastin must be carried out with extreme care, since any error in its ISI will be transmitted to all working preparations calibrated against it and ultimately to large numbers of patients’ INRs.

The calibration of a reference thromboplastin should also be carried out in more than one centre so that (i) the calibration relationship is representative of the different test systems within which the reference preparation will be used, (ii) the suitability of the material as a reference thromboplastin will be comprehensively checked, and (iii) information will be available on the existence and magnitude of possible centre-to-centre variation in the estimates of ISI.

The design of a multi-centre calibration study requires careful statistical planning to ensure that the results will be amenable to a common analysis and to allow valid inter-laboratory comparisons to be made.”

Regarding stability of reference thromboplastins, Kirkwood noted that since these define not a unit of potency, but a calibration relationship, the usual methods for assessing the stability of biological standards are of limited usefulness. Instead a combination of monitoring over time and accelerated degradation tests should be used.

Still, a number of problems remained: “One particular source of remaining difficulty is the specification of normal plasma. While the calibration slope (ISI) of a reference thromboplastin may be measured as identical in two laboratories, if one laboratory restricts (e.g. by age) normals to be members of a group with lower mean coagulability than the other, it would record the same patients’ plasmas as having lower INRs... The ultimate standardisation of the PT ratio will require a consensus as to the definition of normal plasma...

The need for a large number of individual plasmas in a calibration study arises from the scatter of the data points. This scatter is important because it reflects genuine biological variation and because it permits the validity of the assumed straight line relationship to be assessed. It is reasonable to expect pooled plasmas [deep frozen or lyophilised], comprising many individual samples, to be considerably less scattered, so the numbers of tests required to determine the calibration line with a given level of statistical precision could be reduced. However, before pooled plasmas can be used with confidence it must be demonstrated that (i) the calibration relationship between the thromboplastins is satisfactorily linear, and (ii) the pooled plasma gives PTs which lie on or near the straight line that best fits the data points for individual fresh samples. The latter requirement is especially important since it is quite possible that the sample of plasmas which make up the pool are not truly representative of the biological average, or that lyophilisation intro-
duces changes in the coagulation properties of the plasma pool which result in bias of the calibration slope. In the Netherlands, the use of lyophilised patient plasma pools has been shown to be satisfactory for batch to batch calibration of thromboplastins [van den Besselaar et al., 1980] and is being assessed for reliability in the calibration of working reference preparations.”

It is remarkable how these insightful statements by Kirkwood also apply to today’s calibration methodology. It is very probable that many of the discrepancies seen today between INRs determined on the same plasma can be attributed to the fact that only very few of Kirkwood’s recommendations have been implemented.

WHO recommendations, 1983

The recommendations of WHO published in 1983 [43] did not contain recommendations for the calibration of international reference preparations, other than they should “be carried out in international collaborative studies”. Furthermore, the more detailed recommendations for calibration of national or working reference preparations did not at all address the problems of random variation and possible deviations from the assumption of log-linearity. The recommendations provided formulas for the estimation of slopes (with asymptotic confidence intervals) by orthogonal regression, but practically none of Kirkwood’s concerns [42] were specifically addressed.

A collaborative calibration study of reference materials for thromboplastins, 1983

In 1983 J. Hermans et al. published results from a calibration study on six reference thromboplastins performed by seven European and three American laboratories [13]. The six reference thromboplastins to be investigated were WHO’s 67/40 (human), 68/434 (bovine), and 70/178 (rabbit), and three candidate thromboplastins developed in a joint effort of ICSH and BCR: BCT/099 (human brain), OBT/79 (ox/bovine brain), and RBT/79 (rabbit brain). The five thromboplastins were calibrated against the primary IRP 67/40.

“For the seven European laboratories a workshop was organized before starting the study, in order to test the protocol in all details. The American laboratories could not attend this workshop. ... The calibration study was performed in accordance with an extensive protocol. [The detailed protocol can be found in [44].]

To include the effect of interday variation, prothrombin time determinations were performed in each laboratory on at least six, and preferably ten,
different days. On each day each laboratory had to include two freshly pre-
pared ‘normal’ specimens and six freshly prepared patient specimens. Nor-
mals were recruited from the laboratory staff. Patients had to have been
stabilized on anticoagulation; the criterion for ‘stabilized’ was considered to
be a prothrombin time with 67/40 between 1.5 and 5 times the mean ‘normal’
67/40 time for at least 6 weeks. On each day a different set of eight subjects
was taken. All coagulation endpoints were read visually with the tilt-tube
technique. Actual testing was performed such that for each subject all six
thromboplastins were used consecutively; after that the same procedure was
followed for the next subject. The testing order of the subjects was the same
as their order of blood sampling. The order of thromboplastins each day was
varied as prescribed in the protocol.”

The slope of the calibration line (with 67/40 on the vertical axis and the
test thromboplastin on the horizontal) was estimated by orthogonal regres-
sion. The authors noted that “implicit . . . is an assumption that the sta-
tistical errors on both axes are similar in magnitude. This cannot be tested
on our data since they include both biological and experimental variations
and the former cannot be easily reproduced. However, similarity of the two
errors seems reasonable a priori and moreover the result is only marginally
affected if the assumption is not strictly fulfilled.”

A total of 637 plasmas (160 normals and 477 patients) were analyzed in
the 10 laboratories. Of these, (i) 2 normals were discarded due to techni-
cal failures; (ii) 68 patients were discarded owing to their 67/40 time being
outside the interval of 1.5 to 5 times the mean of 67/40 time of the normals
of the corresponding laboratory; and (iii) 3 patient specimens were excluded
because of outlying 67/40 times (outlying being defined as being outside the
region defined by the orthogonal regression line ± 3 times the standard error
around regression, calculated including all observations).”

The authors examined the hypothesis of coinciding regression lines for
normals and patients (log PTs) and found “differences between the two lines
more often than could be expected due to chance.” They showed two plots
where the lines were clearly non-coincident, although the patients’ line in
both cases crossed the data points of the normals. They “felt it justified to
describe the relationship by one line based on the combined patients’ and
normals’ data.”

For each laboratory and each thromboplastin, the slope of the calibration
line was estimated, and for each thromboplastin the ISI was determined as the
unweighted mean of the individual laboratories’ (only the seven European)
estimates of the slopes of the calibration lines.
Tomenson’s modification

Motivated by the finding that the calibration lines corresponding to normals and patients were not necessarily coincident, the statistician J.A. Tomenson in 1984 proposed a modification of the method described by Kirkwood, where the assumption of coincident lines was dropped [45].

“The new calibration method [as described by Kirkwood] has given good results . . . . The main assumption of the model is that the relationship between the logarithms of prothrombin times with different thromboplastins is approximately linear. This has been demonstrated empirically. An additional finding has been that the use of logarithms also achieves homoscedasticity. However, the new calibration model does have one important weakness that was apparent in both calibration studies and particularly in the study by van den Besselaar and van der Velde [46] in which manufacturers calibrate their products against the BCR reference materials. This is the failure in several instances of the normal control plasmas to lie on the line through patients’ plasmas. . . .

If the same calibration equation is valid for patients and normals it can be shown [42] that the calibration model implies a relationship between prothrombin time ratios of the form

$$r_1 = r_2^b \quad (1)$$

where $r_1$ and $r_2$ are prothrombin time ratios with two thromboplastins and $b$ the slope of their calibration line. The success of the new calibration model hinges on the ability of this model to describe the relationship between prothrombin time ratios. It should come as no surprise after previous experience with the Biggs-Denson procedure that this model will occasionally prove unsatisfactory. . . . Just as the calibration lines of the Biggs-Denson procedure were forced to pass through the point (1,1) so the intercepts of the [lines corresponding to equation (1)] are also restricted. The natural way to overcome this problem is to introduce a scale parameter and use a model for prothrombin time ratios of the form

$$r_1 = 10^d r_2^b \quad (2)$$

Clearly equation (1) is a particular example of equation (2) but the new model will also cope with data sets for which the mean logarithms of the prothrombin times of normals do not lie on the orthogonal regression line of the patients.

The estimation of $d$ is made easier if a further slight modification is made to the calibration model. This is merely an assumption that the prothrombin
times of normal control plasmas should each follow the same distribution rather than having separate individual distributions as is currently assumed. With this assumption it can be shown that $d$ is estimated as

$$d = \bar{y}_n - a - b \bar{x}_n \quad (3)$$

where $\bar{x}_n$, $\bar{y}_n$ are the mean logarithms of the prothrombin times of normals and $a$, $b$ the slope and intercept of the orthogonal regression line calculated using only the results of patients. . . .

It is not reasonable to expect the plasmas of normal controls [and patients’ plasmas] to follow the same relationship since there are major biological differences between patient and normal plasmas. More importantly the relationship observed for normal plasmas may simply result from correlation between the measurement errors [see figure 3 in [45] for a good illustration of this]. . . .

For equation (1) to be used in a conversion formula for prothrombin ratios it is sufficient to demonstrate that the mean logarithms of the prothrombin times of normals lie on the regression line through the patients’ plasmas. A test of this hypothesis can alternatively be regarded as a test that $d = 0$ in equation (2). The use of equation (2) enables a more logical approach to be made to calibration of thromboplastins. The first step is to fit a model of the form specified by equation (2) which amounts to deriving the orthogonal regression line for patients’ plasmas and estimating $d$ by use of equation (3). A test can then be made of the hypothesis that $d = 0$ and an approximate test of this hypothesis is given in the appendix [in [45]. See also page 50 in this dissertation]. If the scale parameter is not significantly different from unity, the simpler equation (1) may be used and calibration proceeds in the usual manner. . . .

The modification of the calibration procedure is in essence little more than a formal rendering of the pragmatic approach adopted in [46]. The results of the European (BCR) calibration exercise, the international collaborative study and the Manufacturers’ calibration study . . . suggest that the generalised model will rarely be needed. For most calibrations the mean logarithms of the prothrombin times of normals do lie on the orthogonal regression line through patients’ plasmas. Nevertheless it is important that clear guidelines should be available in the event that the assumptions of the calibration model are not satisfied.”

In a trial from 1986 Gogstad et al. [47] provided an example, where the correction suggested by Tomenson was very appropriate: “The combined thromboplastin reagent, Normotest, has been calibrated against the secondary international reference preparation for bovine thromboplastin, OBT/79.
Three expert laboratories measured up to 62 patients on stabilized oral anticoagulant therapy and up to 20 normals in order to establish an INR-scale for Normotest. It was found that the model recommended by the WHO was less suited for the calibration of this thromboplastin. This is the first study in which three independent laboratories demonstrate a similar bias of the WHO calibration model. A modified model in which a correction factor is introduced was applied to the problem and proved to give a reliable calculation method for INR on Normotest. The mean coefficient of variation of INR calculated between measurements with Normotest and OBT/79 (scatter of data around calibration line) was 4.2–5.0% as compared to 5.1–5.7% for the WHO-method. A conversion scale for percent activities between Normotest and Thrombotest was established showing that the recommended therapeutic range of 5-10% Thrombotest (INR = 4.8–2.8) corresponds to 10–20% Normotest.”

A recent study [48] found that the use of Tomenson’s correction could reduce the deviation between assigned and locally determined INR values of lyophilized plasmas.

Formulae for orthogonal regression were also provided by van der Velde [49], who also provided a test for the coincidence of the orthogonal regression lines of normals and patients. This test seems, however, obsolete in the light of the less strict requirement of Tomenson, where it suffices that the line of the patients passes through the point corresponding to the mean normal log PTs.

Uncertainty of the estimates

Tomenson also considered the question of the appropriate inference space: “The uncertainty in an estimate of the slope of the calibration line and the standard deviation about it are dependent on the use to which it is put. Inter-laboratory differences in calibration relationships must be considered if a calibration line is to be used outside the laboratory where it was derived.

Manufacturers are recommended to calibrate a house standard against the BCR rabbit thromboplastin RBT/79 and to calibrate production lots against the house standard. The cumulative inaccuracy for such a calibration is not known but it will depend on the correlation between the two calibration steps involved in the calibration of a production lot against the BCR rabbit thromboplastin. The BCR study has shown that there is a strong correlation between the results of different calibrations performed within the same laboratory.

Hermans et al. [13] report the slopes of the orthogonal regression lines
2. BACKGROUND

for the calibration of the BCR reference materials and the secondary WHO IRPs against the primary IRP obtained by each of the ten laboratories in the BCR study. The correlation between the calibration lines of the WHO and BCR rabbit thromboplastins is 0.78 and 0.89 for the WHO and BCR bovine thromboplastins. If a CV of 5% is assumed for each of the two calibration steps, house standard against production lot and house standard against the BCR rabbit thromboplastin, a CV of between 7 and 10 per cent will result for the combined calibration. This figure is dependent on the strength of the correlation between the two calibrations and the results of the BCR study indicate a CV much closer to 10 per cent than 7 per cent. There is also uncertainty in the third step of the calibration, the calibration of the BCR rabbit thromboplastin against the IRP, and this will increase the total CV of the calibration to 10.5 per cent. The uncertainty in an INR is also dependent on the variability of points about the calibration line. Again the magnitude of the standard deviation about the orthogonal regression line for a calibration of a production lot of rabbit thromboplastin against the IRP is not known. Calibration of the BCR rabbit thromboplastin against the IRP would suggest a figure not less than 0.03.”

Under slightly simplifying assumptions, Tomenson found that for an INR estimated at 3.5, the 95% confidence interval is as large as 2.47 - 4.96 (page 98 in [45]). “This is a wide interval and it is clear that some patients whose INR is well within the therapeutic range could actually be under-dosed or over-dosed. Nevertheless this is probably the worst case likely to occur in practice and the position could be considerably improved if manufacturers obtained independent estimates of the calibration of the house standard against the BCR rabbit thromboplastin.

For conversion of prothrombin ratios a further component should be included in the standard error of prediction to account for the variation in the mean normal prothrombin time. This term is of less importance as it is given by $s^2/n$, where $s^2$ is the average variation of observations about the calibration line and $n$ the number of normals used to calculate the prothrombin time ratio.”

These problems are addressed in much more detail in chapter 5 of this dissertation.

Applicability of INR to patients in early stages of therapy

With regard to the use of the INR in monitoring patients in early stages of therapy, Tomenson [45] had the following comments: “Loeliger and Lewis . . . have proposed ranges expressed as INRs and as ratios with the BCR reference materials for various clinical conditions for both in-patients and out-
patients. However, the protocol of the BCR study specifies that patients used for calibration should have recovered their general health (preferably outpatients) and have been stabilised on anticoagulation for at least six weeks. The theory underlying anticoagulation would suggest there is no reason to assume that certified values obtained using such patients can be used to calculate INRs for patients in the early stages of therapy. Coumarin-type anticoagulants induce a combined depression of the coagulation factors II, VII, IX, and X with factor VII being reduced first and factor II more slowly. Individual thromboplastin reagents vary considerably in their sensitivity to the factors II, VII, and X and it might be anticipated that different calibration results would be achieved for patients in the early stages."

To investigate this hypothesis, Tomenson performed a small exercise with two sets of calibrations, one for stabilized patients and one for patients who had undergone less than three weeks of therapy. “The principal difference between the two series of calibrations was found in the standard deviations about the calibration line. . . . It can be seen that the variation about the calibration line is greater for short-term patients. In other words the procedure is less dependable in non-stabilized patients. The effect was much more marked for those patients seen during the first five days of therapy although the numbers were too few to draw firm conclusions.”

Species-to-species or like-to-like?

On the choice of calibration sequence, Tomenson [45] had the following comments: “The BCR calibration protocol recommends that [manufacturers] use the reference of the same species [as their own product] because it is felt that for this calibration the variation of prothrombin times about the line is the least. In the BCR study this was the case for two of the three BCR reference materials when calibrated against their WHO counterparts. However, re-analysis of the BCR data has shown that the principal factor influencing the variability of prothrombin times about the line is not the species effect i.e. whether of human or animal origin, but whether the like-to-like relationship related to the composition of the reagent i.e. whether it is the usual Quick test reagent [i.e. plain] or manufactured with added plasma [see the splendid figures 6-8 in [45]]. . . .

For the IRP 67/40, which is a human combined reagent, the best calibration is against the bovine combined reagents. However, amongst the plain reagents, the best calibration is against the BCR human brain reagent BCT/099. The bovine WHO reference preparation, a combined reagent, calibrates best against the BCR bovine thromboplastin. In the case of 70/178 a plain, rabbit reagent, the best calibration is achieved against the BCR human
and rabbit thromboplastins, both plain reagents. The calibration against the BCR rabbit thromboplastin is only marginally better than against the BCR human thromboplastin and would make little practical difference. The Quick prothrombin time test [i.e. using plain thromboplastin] is the procedure used in the hospitals of most countries for anticoagulant control [Lam-Po Tang & L. Poller, 1975]. This result strengthens the case put forward by Ingram [Ingram, 1979] for the replacement of the WHO IRP [67/40] which is a combined reagent by the proposed second IRP which is a human brain Quick test (plain) reagent.”

**Further improvements and discoveries on INR**

Since the introduction of the INR standard an overwhelmingly large amount of literature has been published on this subject. It is impossible to provide a fair overview of all these contributions in the frames of this dissertation, and in the following I shall only draw attention to a few of those with immediate relevance to the three substudies in this dissertation.

**Calibrations of international reference preparations of thromboplastin**

The calibrations of IRPs have been described in a series of publications, which have provided the details that were used for setting up the general calibration model in chapter 5. The first calibration exercise [13] has already been described in details on page 25 ff. The following briefly summarizes the published information on the IRPs that have been calibrated so far. In all cases the WHO calibration model was found to be adequate, such that the relation to IRP 67/40 could be sufficiently expressed by the ISI. Note, that many of the IRPs have expired, since the supply of them has been exhausted. The sources of information as well as the reported ISI estimates and standard errors, se(ISI), of these are presented. Moreover, the reported common slope estimate, b, the between-laboratory standard deviation of the slope estimate, sd(b), and the number of participating laboratories, n, are presented.

**67/40** Human, combined [37]. With bovine factor V and fibrinogen. Reconstituted with 2 ml of 3.2 mM CaCl₂ and used in 0.4 ml amounts with 0.05 ml of plasma [27]. The first IRP. ISI = 1 per definition.

**68/434** Bovine, combined [37]. With bovine factor V, fibrinogen and cephalin. Reconstituted with 2.2 ml of 3.2 mM CaCl₂ and used in 0.4 ml volumes with 0.05 ml of plasma [27]. Calibrated against 67/40. ISI = b = 1.027, n = 7, sd(b) = 0.042. No se(ISI) was provided [13].
2.1. PLASMA BASED METHODS

69/223 Human (brain) [37]. Freeze-dried phenolized extract of human brain. Reconstituted with 2 ml distilled water and used in 0.1 ml amounts with 0.1 ml of plasma and 0.1 ml of 25 mM CaCl₂ [27].

70/115 Rabbit, combined. Reconstituted with 2.2 ml of distilled water and used in 0.25 ml volumes with 0.01 ml of plasma [27].

70/178 Rabbit, brain (plain, according to [13]). Freeze-dried phenolized extract of rabbit brain. Reconstituted with 1 ml distilled water and used in 0.1 ml volumes with 0.1 ml of plasma and 0.1 ml of CaCl₂ [27]. Calibrated against 67/40. ISI = b = 1.617, n = 7, sd(b) = 0.074. No se(ISI) was provided [13].

70/183 Rabbit (brain) [37].

BCT/099 Human, plain (brain). A lyophilized batch of British Comparative Thromboplastin, prepared courtesy of Dr. L. Poller, Manchester, UK. To be reconstituted with 1.0 ml water + phenol (0.5 g/l). pH = 6.5. Calibrated against 67/40. ISI = b = 1.048, n = 7, sd(b) = 0.040. No se(ISI) was provided [13].

OBT/79 Bovine, combined (brain). Prepared courtesy of Nyegaard & Co., Oslo. To be reconstituted with 2.2 ml of CaCl₂ (3.2 mM). pH = 7.73. Calibrated against 67/40. ISI = b = 1.011, n = 7, sd(b) = 0.040. No se(ISI) was provided [13].

RBT/79 Rabbit, plain (brain). Prepared courtesy of Dr. K.W.E. Denson, Thame UK. To be reconstituted with 0.5 ml distilled water. pH = 6.8. Calibrated against 67/40. ISI = b = 1.413, n = 7, sd(b) = 0.096. No se(ISI) was provided [13].

BCT/253 Human, plain (brain). To be reconstituted with 0.5 ml phenolized water (250 mg/l). Calibrated against IRP 67/40. ISI = b = 1.1, n = 16, sd(b) = 0.054. No se(ISI) was provided. [50]. This ISI is biased, according to van den Besselaar [17].

BCT/441 Human, plain (brain). To be reconstituted with 1.0 ml of phenol solution (0.175 %). pH = 6.3. Calibrated against BCT/253. ISI = 1.041, n = 7, b = 0.959, sd(b) = 0.038. No se(ISI) was provided [15]. This ISI is biased (according to van den Besselaar), since it is calibrated against BCT/253.

CRM 149R Rabbit, plain. Calibrated against RBT/79. ISI = 1.343, se(ISI) = 0.035, n = 11, b = 0.950, sd(b) = 0.019 [16].
2. BACKGROUND

**RBT/90** Rabbit, plain. Calibrated against RBT/79. ISI = 1.035 (recommended to round off to ISI = 1.0), se(ISI) = 0.027, n = 20. b and sd(b) were not provided [17].

**CRM 149S** Rabbit, plain. Calibrated against RBT/90. ISI = b = 1.257, se(ISI) = 0.013, n = 11. sd(b) was not provided [18].

**X/95** Human, recombinant. Calibrated against BCT/253, RBT/90, OBT/79. ISI = 0.940, se(ISI) = 0.0060 [19].

**Accuracy of INR estimates**

The imprecision of INR is composed of the within- and between-laboratory variation. According to Loeliger et al. [51] the between-laboratory variation of INR was approximately 7–8% in the Netherlands and in Britain in 1985, and the overall coefficient of variation (CV) of the INR was 11–13.5% under well-controlled conditions.

Other studies [52, 53, 54] reported between-laboratory CVs ranging from 1% to 7%. Many studies have investigated the effect of instrumentation on ISI [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66]. All of these found that the instrument has an effect on the ISI and it is recommended that every local PT system should be assigned its own ISI by local calibration. One study [67] of laboratories using the same type of steel ball coagulometer (KC) found practically no difference between ISI estimated by automatic and manual method. The clotting times, however, were shorter by the automatic method.

Discrepant INR values are reported from time to time [68, 69, 9, 70, 71, 72, 11]. This is not so strange, since an infinite number of factors may influence the INR. Some of the factors that have been considered are exposure to carbon dioxide [73], volume errors [74], and concentration of citrate, where an increase in concentration from 109 mM to 129 mM may reduce the ISI by 10% [75, 76]. Moreover INR may not be reliable in the early phases of OAT or in unstable periods [77, 78].

A number of studies have indicated that the discrepant INR values are not caused by degradation of reference thromboplastins [79, 80], nor by freezing or storing at 2–8 degrees C for up to 24 hours [81, 82]. Two studies showed that the use of two tubes for blood sampling is superfluous.

**Local calibration**

Although this thesis is primarily concerned with the calibration of IRPs, a few words shall be said on local calibration. Recently a large number of pub-
2.1. PLASMA BASED METHODS

Applications have presented studies on the use of lyophilized or frozen, artificially depleted or coumarin plasma for ISI assignment in local laboratories. The recommendations are divergent. Some studies found no appreciable difference between fresh and freeze-dried plasma \[83, 84, 85, 86, 87, 88\]. Others found a definite prolongation of PT by freezing and especially freeze-drying \[89, 90\]. Most agree, however, that the use of freeze-dried or frozen plasmas, at least under certain conditions, may reduce the inter-laboratory variance \[91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101\].

**Latest WHO recommendations (1999)**

The latest WHO guidelines for thromboplastins \[102\] are based on the same empirical observations as those from 1983. They are, however, more detailed. Among the most important additions are:

- Methods are recommended for the calibration of IRPs.
- ISI is defined for PT systems, i.e. instrumentation and thromboplastin combined.
- The definition of INR has been extended to cover whole blood.
- A number of tests on thromboplastins have been introduced to ensure more uniformity.
- A distinction is made between calibration of a) IRPs; b) secondary standards (national reference preparations and manufacturer’s working standards); c) commercial preparations; and d) local systems.

The guidelines contain the following important definitions:

**Prothrombin time (PT)** The clotting time of a plasma (or whole blood) sample in the presence of a preparation of thromboplastin and the appropriate amount of calcium ions. The time is reported in seconds.

**Prothrombin-time system (PT system)** A procedure by which the PT is determined using a specific thromboplastin reagent and a particular method, which may be manual, e.g. a tilt-tube method, or involve the use of an instrument that records the coagulation endpoint automatically. The method should be described and the description should include all procedures and equipment used, e.g. the pipettes and tubes.

**Mean normal prothrombin time (MNPT)** The geometric mean of the PTs of the healthy adult population. For practical purposes, the geometric mean of the PTs calculated from at least 20 fresh samples
from healthy individuals, including those of both sexes, is a reliable approximation of MNPT. It is not necessary to collect and test all the individual samples in one session. It is recommended that each laboratory should determine MNPT using its own PT system. Pooled normal plasma (either deep-frozen or freeze-dried) may be suitable if the clotting time obtained is related to the MNPT value and its storage stability is acceptable.

**Prothrombin-time ratio (PT ratio)** The PT obtained with a test plasma or whole blood divided by the MNPT, all times having been determined using the same PT system.

**International Sensitivity Index (ISI)** A quantitative measure, in terms of the first IRP 67/40, of the responsiveness of a PT system to the defect induced by oral anticoagulants.

**International Normalized Ratio (INR)** For a given plasma or whole blood specimen from a patient on long-term OAT, a value calculated from the PT ratio using a PT system with a known ISI according to the formula

\[ INR = \left(\frac{PT}{MNPT}\right)^{\text{ISI}}. \]

### 2.1.5 Danish quality controlled laboratories

Most Danish hospital laboratories participate in a quality assurance scheme organized by the Danish Institute for External Quality Assurance in Hospital Laboratories (DEKS). Each local PT system (comprising the thromboplastin preparation and the instrumentation), is assigned an ISI value based on a two-point calibration using two frozen plasma samples supplied by DEKS. The “ISI-calibrator” consists of pooled plasma from 20 patients on stable OAT, and the “normal calibrator” is pooled plasma from 40 healthy individuals. The ISI calibrator is assigned an INR value by DEKS, \(\text{INR}_{\text{abnormal}}\), using an IRP. In the local PT system, the geometric mean PT of the ISI calibrator, \(\text{PT}_{\text{abnormal}}\), and the normal calibrator, \(\text{PT}_{\text{normal}}\), are obtained by repeated PT determinations using the standard operation procedure (SOP) of the laboratory. These are used to assign an ISI value by the formula

\[ \text{ISI}_{\text{local}} = \log(\text{INR}_{\text{abnormal}}) / \log \left( \frac{\text{PT}_{\text{abnormal}}}{\text{PT}_{\text{normal}}} \right). \]

Subsequently patient INRs are routinely calculated by the formula

\[ \text{INR} = \left( \frac{\text{PT}}{\text{PT}_{\text{normal}}} \right)^{\text{ISI}_{\text{local}}} \]

where PT is obtained by the SOP.
2.2 Capillary whole blood coagulometers

Portable capillary whole blood coagulometers, which do not require centrifugation of or addition of citrate to blood samples, are well suited for patient self-management of OAT if they can measure INR accurately. A number of studies have sought to investigate the accuracy of such coagulometers. Most of these, however, compared with routine laboratory methods and are thus of limited use with respect to trueness. Moreover, precision was often inadequately evaluated, since only measures of correlation are presented.

In the following some studies are presented, which investigated two essentially identical coagulometers.

**Coumatrak Protime**

Coumatrak Protime (also called Biotrack 512 [103, 104], Protime [105], Protime 1000 [106], and CoaguChek Plus [107]) was approved by the American Food and Drug Administration (FDA) for home use by patients in 1997. It was described by Lucas et al. [108] in 1987: “Recently, a capillary whole blood system (Protime Monitor 1000, Biotrack Inc., Mountain View, CA) was developed for measuring PT that overcomes many of the disadvantages of previous methods. . . .

The Protime Monitor 1000 is a laser photometer that is portable (5” × 9” × 2”) and battery powered. Before application of a blood sample, a disposable plastic reagent cartridge is inserted into the monitor for warming to 37 degrees C. When a drop of whole blood (minimum of 25 ml) is applied to the cartridge, it is drawn by capillary action into a reagent chamber that contains a dry rabbit brain thromboplastin preparation. Coagulation begins when the blood rehydrates the thromboplastin. The reaction mixture continues flowing in the capillary channel beyond the reagent chamber before a clot develops. The laser photometer detects the cessation of blood flow (clotting) by sensing variation in light scatter caused by the movement of red blood cells. The time elapsed between application of blood and cessation of flow is measured by the monitor and is converted mathematically to a plasma equivalent PT. The conversion factors between whole blood and plasma PTs were generated from a large population of normal and anticoagulated subjects. The plasma PTs were obtained using rabbit brain thromboplastin (Thromboplastin C, American Dade Inc. Miami, FL) and an MLA 750 instrument (Medical Laboratory Automation Inc., Mount Vernon, NY).

After a PT is displayed, a new cartridge may be inserted for the next test. Each test requires approximately 1 minute.”

Lucas et al. performed the following substudies:
1. Comparison of Protime with plasma method (Dade C): Unfortunately only correlations (between PTs of Protime and plasma methods) were considered. The authors concluded that “the correlation coefficient of 0.96 indicates good overall agreement”, but the authors did not further elaborate on this. The plot of Protime whole blood PT against reference plasma PT (figure 2 in [108]) indicates a curvilinear relationship such that normals tend to lie below the orthogonal regression line. The presented data do not provide a basis for evaluating the accuracy of Protime.

2. Comparison of capillary and venous whole blood on the Protime System: The results of the study are given as means and standard deviations for each method, and correlations between methods. These results do not allow for evaluation of differences between whole blood and plasma.

3. Precision of the Protime System: “The within-day precision of the capillary PT was evaluated ... by measuring whole blood controls ... ten times each on two different Protime monitors. The two instruments gave similar results so the data were combined. A new vial of control was used for each pair of determinations. The control material was formulated by mixing fixed human red blood cells with either normal plasma or plasma made deficient in heat-labile coagulation factors. ... The use of Control I gave a mean and CV of 12.3 seconds and 4.9%. The mean and CV for control II were 20.0 seconds and 2.9%. ... The between-day precision was evaluated at all centers by performing measurements for Control I and Control II over the six-month period of the study. ... The CV for Control I varied from 3.5 to 5.0% and for Control II from 2.2 to 7.4%. ...

Within-day replication of capillary times was evaluated at all centers by testing two separate fingerstick samples from 137 subjects (normal and anticoagulated) on two separate monitors. ... The correlation coefficient was 0.99, and the slope of the regression line was 1.00.” Unfortunately, the precision of measurements on patients’ blood samples was not evaluated.

4. Effect of hematocrit on the Protime System: “Blood anticoagulated with EDTA was obtained from 393 subjects ... . The ratio of capillary PT to the reference time was examined as a function of hematocrit. ... [Hematocrit] ranged from 23.4% to 53.8%, and the mean (± SD) was 40.6% ± 5.4%. ... Over this range there was no effect of hematocrit
on the ratio of capillary PT to the reference PT (correlation coefficient $r = 0.08$, $p > 0.05$).”

The authors concluded that “the capillary PT device described in this report is a technical advance that provides immediate results, obviates the need for venipuncture, and appears to be comparable in accuracy to current reference methods.”

Weibert & Adler [109] found a bias of 7% between routine laboratory methods and Protimé. The precision was not explicitly evaluated. The authors concluded that “the Coumatrak system can be recommended for routine use in monitoring outpatient anticoagulant therapy.”

Belsey et al. [110] evaluated precision by testing normal and abnormal artificial control plasmas in duplicate twice each study day for 20 days. They found a CV for PTs of normal and abnormal controls of 4–6% and 8–10%, respectively. The bias between monitor and laboratory measurements was 10%. They concluded that “the Coumatrak can rapidly provide PT test results that are clinically useful for the office management of patients being treated with a warfarin anticoagulant and for the diagnosis of selected disorders. The system was found to be easy to operate, appropriate for use by individuals with little laboratory experience, and was subject to few operational problems during this study.”

The use of artificial control plasmas may cause the imprecision to be overestimated, since the coagulometers are optimized for testing whole blood. Jennings et al. [103] also examined precision by means of artificial control plasmas. They found that CV of INRs by low- and high-level control plasmas were 7.5% and 4.5%, respectively. By comparison with routine laboratory methods they found that the bias between monitor and consensus INR, routine laboratory and consensus INR, and monitor and routine laboratory INR were -0.35 INR, 0.44 INR, and -0.76 INR, respectively. They argue that the manufacturers should allow for local calibration of the coagulometers.

McCurdy & White [105] estimated the precision by measuring INR on repeat samples for each patient, taken within two minutes. They found that the CV for the monitor was 7.7% and that the similar value for the routine laboratory was 6.3%. Moreover they noted that, “the monitor measured on average 0.3 INR units high at a criterion [i.e. laboratory] value of 2.0 units and less than 0.1 units low at a criterion standard value of 3.0 INR units. At a criterion standard value of 4.5 INR units, the monitor measured on average 0.5 units low.” This observation is, however, a statistical artifact, which will be seen when laboratory INR is measured with error, even though both laboratory and monitor INRs are unbiased for the true INR. Thus it cannot be concluded that the monitor underestimated INR for high levels of
InR. For a detailed discussion of this, see page 95.

Tripodi et al. [111] performed a calibration exercise, where the monitor was calibrated against CRM 149R according to WHO recommendations. By using artificial control plasma they found that the CV of the monitor was 9.7% for PTs and 18.8% for INRs. The similar figures for CRM 149R were 4.8% and 6.5%. They also found that the monitor INR, calculated from monitor PT by using the manufacturer’s ISI value and MNPT, agreed poorly with CRM 149R. By using the ISI value obtained by the calibration, the agreement was better. However, the most important result from the study was that “the monitor can be calibrated with the WHO model, because log-transformed PTs for patients stabilized on oral anticoagulants and normal individuals are linearly related and because the same orthogonal regression line describes patient and normal data points adequately”.

Anderson et al. [106] did not evaluate precision. They found that the monitor agreed with standard criteria for 83% of the INR measurements.

Kaatz et al. [4], by using venous blood, found CVs for Coumatrak, routine laboratory and manual tilt tube of 4.18%, 1.5% and 1.14%, respectively.

CoaguChek

The CoaguChek (Roche Diagnostics) coagulometer works in essentially the same way as Protime, except that clotting is detected photometrically by the cessation of the movement of iron particles in a magnetic field in the test strip. The CoaguChek System consists of the CoaguChek coagulometer, CoaguChek PT Test test carriers, a code chip specific to the batch of test carriers, CoaguChek PT Controls control plasma, and a Softclix lancing device for skin puncture for collection of capillary blood. The code chip contains a calibration curve, obtained by the manufacturer by calibration with citrated venous plasma of samples from the given batch of test carriers against the Hepato Quick method (Roche Diagnostics).

The coagulometer is operated by inserting a test carrier into the meter. After a series of self-check procedures, including a check of consistency of the code chip with the test carrier, the meter is ready for analysis. A tip of a finger is punctured with the lancing device, and a drop of capillary blood is applied onto the test carrier’s application zone. After approximately 1 minute, the test result is displayed as INR. The performance of the coagulometer can be checked by the patient or other operators using a control plasma supplied by the manufacturer. The previous 30 test results are saved in the memory of the coagulometer specifying INR value, date and time of measurement, allowing for instant inspection.

The precision of CoaguChek was also investigated by Kaatz et al. [4].
They found a CV of 3.14%.

van den Besselaar et al. [112] examined both trueness and precision of INR measurements by CoaguChek. The within-run precision using artificial controls and venous blood was 6–7% and 3–4%, respectively. The authors noted that “the convincing precision results for blood are due to the fact that the CoaguChek system has been optimized for this sample material.” By comparison with other reagents they found a bias between monitor vs. Hepato Quick, monitor vs. Thromborel, monitor vs. CRM 149R, and Hepato Quick vs. Thromborel of 0.21 INR, 0.05 INR, 0.05 INR, and 0.16 INR, respectively. The authors concluded that “its good analytical performance and convenient handling recommend the CoaguChek system as a suitable system for decentralized prothrombin time testing.”

Douketis et al. [113] found that 79% of 163 dual measurements by CoaguChek and a laboratory-based method were within 0.5 INR units. This proportion depended on the level of laboratory INR, since the proportion of dual measurements within 0.5 INR units for laboratory INR ranges of < 2.0, 2.0–3.0, 3.1–4.0 and > 4.0 was 98%, 87%, 57%, and 21%, respectively. The decline in proportions within a fixed range for increasing INR is consistent with the general finding that the standard deviation of INR measurements is proportional to the level of INR. Douketis et al. also found that CoaguChek underestimation of INR increased with increasing levels of laboratory INR. As noted above this is a statistical artifact and it cannot be concluded that the monitor underestimated INR for high levels of INR.

CoaguChek was also evaluated by Hasenkam et al. [114]. Twenty patients on OAT had simultaneous blood samples taken from two finger punctures by a trained laboratory technologist. These samples were immediately analyzed on two CoaguChek instruments. Out of 36 successful double determinations, 28 were within ± 0.1 INR and the remaining within ± 0.3 INR. In a study of 20 patients on self-managing OAT, the median difference between routine laboratory INRs and monitor INRs measured at home (191 double determinations) was 8.0% [115]. The CV of the differences between the two methods was 15.6%, with 8.9% between patients and 12.8% within patients. The CV within patients could be attributed to long-term analytical variation in the monitor (10%) and in the laboratory (8%). However, since the within-patient CV also comprises biological variation in the time between the two measurements, the true values of the analytical CVs are probably lower. The CV between patients was caused by patient-specific differences, ranging from –11% to 21% [116]. The patient-specific differences may be device-specific or due to interactions between measurement method (capillary blood or plasma) and patients.
3

Materials and methods of empirical studies

3.1  Precision of home coagulometer

3.1.1  Study subjects

Thirteen patients (2 females and 11 males; median age 63 years; range 43 to 71 years) who had performed self-management of anticoagulant therapy using a CoaguChek home coagulometer for at least six months agreed to participate in the study. The clinical performance of a subset of 11 of these patients has been published elsewhere [114, 116]. Their median time with completed full-scale self-management was 591 days (range 182 to 632 days). OAT consisted of Phenprocoumon (Marcoumar) in all cases, aiming at a therapeutic range of 2.0-3.0 INR except for two patients (nos. 14 and 19), having a target range of 2.5-3.5 INR.

3.1.2  Patient self-testing of INR by the home coagulometer

The procedure for patient self-testing has been described in details on page 40.

3.1.3  Laboratory measurements of INR

Citrated venous platelet-poor plasma was obtained by cubital venipuncture and analyzed at Skejby Sygehus both in the beginning and end of the study, and at Roskilde Amtssygehus in the end of the study. Samples for analysis by Roskilde Amtssygehus were frozen and sent overnight in a container with dry ice.
Both laboratories participated in DEKS’ quality assurance scheme (see page 36). The ISI-calibrator used at Skejby Sygehus was the DEKS ’94 calibrator, which was assigned an INR value using RBT/90. The ISI-calibrator used at Roskilde Amtssygehus was the DEKS ’96 calibrator. This was assigned an INR value using OBT/79 (bovine, combined) and BCT/253 (human, plain) with the manual tilt tube technique.

The analyses at Skejby Sygehus were performed according to the following standard operation procedure (SOP): For each plasma sample, 85 µl Nycotest buffer is pipetted into a cuvette and thoroughly mixed with 15 µl citrated plasma. The procedure is repeated with another cuvette and then the two cuvettes are inserted into a carousel with space for 12 cuvettes. This is inserted into the coagulometer (Coag-A-Mate X-C, Organon Teknika, Turnhout, Belgium) and heated to 37 C. Subsequently the coagulometer determines the PT for each cuvette by injecting 200 µl Nycotest PT reagent (Nycomed Pharma, Oslo, Norway) and registering the time until clot formation. The sample PT is calculated as the mean of the two measurements.

Analyses at Roskilde Amtssygehus were performed using a Thrombolyzer coagulometer (BE Behnk Elektronik, Hamburg, Germany) and the SPA 50 (Diagnostica Stago S.A., Asnieres, France) reagent.

Both of the reagents used in the study are combined thromboplastin preparations, demonstrating specific sensitivity to coagulation factors II, VII and X.

### 3.1.4 Study design

Patients were instructed to continue self-management of OAT as previously, but to perform an additional INR measurement on their coagulometer immediately after the usual weekly home measurement. The second test was done using a new test carrier and blood from another finger, and both INR determinations were recorded by the patient on a dedicated form. To document all measurements performed on the patient’s coagulometer, any extra measurements should be entered as well. This procedure was consistently followed during the study period of 10 weeks (June to August 1997) for all patients. All patients used the same batch of test carriers (lot 44) throughout the study period.

On the same day (10. June 1997) all patients had two blood samples taken from cubital venipuncture, and both blood specimens were analyzed by routine laboratory methods at Skejby Sygehus. At the same time (within the following hour) the patient performed the first double measurement on his/her coagulometer.

In the end of the study period (August 1997) all patients were once again
seen at Skejby Sygehus for a blood sample from cubital venipuncture. The sample was prepared for analysis and then split into two portions of which the first was analyzed by Skejby Sygehus and the second was frozen and stored at -80°C. Subsequently (within the next hour) each patient performed two sets of double measurements on his/her coagulometer: one pair of measurements using lot 44 of test carriers and another pair using a different batch (lot 45). On the same occasion, the patient handed over the completed form with results from his/her self-testings. Moreover, the results of all 30 previous measurements stored in the memory of the coagulometer were downloaded to a worksheet. A few days after all patients had been seen, frozen samples were analyzed on the same day by Roskilde Amtssygehus.

### 3.1.5 Statistical methods

To stabilize variances and make distributions more normal, the statistical analyses were performed on natural logarithmically transformed INR values.

To discriminate between overlapping data points in plots, a jittered technique was used in plotting where appropriate.

Statistical analyses were carried out using the SAS/STAT (SAS Institute, Cary, North Carolina, USA) and the S-Plus (MathSoft Inc., Seattle, Washington, USA) statistical software packages. Long-term within-patient biological variation and long-term analytical variation was estimated in a General Linear Mixed Model using PROC MIXED in the SAS/STAT package.

### 3.2 Validity of the INR standard

#### 3.2.1 Plasma samples

Abnormal plasma samples were obtained from 55 consecutive patients on stable long-term oral anticoagulant therapy (phenprocoumon), who visited Skejby Sygehus in November or December 1997. Samples were taken by cubital vein puncture in a 5 ml Venoject tube containing 3.2% sodium citrate. After centrifugation, 500 µl of each sample was pipetted into each of three NUNC Cryotubes, 1.8 ml, then immediately frozen in liquid nitrogen (-177°C), and stored at -80°C until analysis.

Normal plasma samples (frozen and stored on dry ice) were obtained from 10 healthy donors (non-smokers, non-obese, less than 50 years of age) in Indianapolis, USA, by James L. Hill.
3. MATERIALS AND METHODS OF EMPRICAL STUDIES

3.2.2 Measurement methods

Nycotest

The ISI-calibrator used at Skejby Sygehus was the DEKS '96 calibrator. This was assigned an INR value of 2.15 by A.M.H.P. van den Besselaar, Leids Universitair Medisch Centrum, The Netherlands, using OBT/79 (bovine, combined) and BCT/253 (human, plain) with the manual tilt tube technique.

Nycotest (Nycomed, Oslo, Norway) is a rabbit, combined thromboplastin preparation. PT determinations by Nycotest were performed at the Department of Clinical Biochemistry, Skejby Sygehus according to its SOP as specified on page 44.

Two weeks before the PT determinations on frozen plasma, in February 1998, the local laboratory with its SOP was assigned new ISI and MNPT values. The local MNPT was obtained as the mean of 24 PT determinations of DEKS' normal plasma calibrator (mean 20.05 sec, standard error of the mean, se, 0.15 sec).

Hepato Quick

Hepato Quick (Roche Diagnostics, Mannheim, Germany) is a rabbit, combined thromboplastin preparation. Lot 667017 with an assigned ISI of 0.99 was used. The assigned MNPT was 21.05s. PT determinations were performed on a KC4A instrument (Amelung, Germany) according to the following procedure. The coagulometer contains four cups. For each plasma sample, a steel ball is added to a cup followed by 10 μl citrated plasma. Next, 200 μl Hepato Quick reagent is pipetted into the cup. The procedure is repeated with a new cup and plasma from the same sample. The cups are allowed to incubate 2 minutes before adding 100 μl of prewarmed 10 mM calcium chloride. At the same time the test begins and the time until clot formation is registered. The sample PT is calculated as the geometric mean of the two measurements.

CRM 149S

PT was determined by manual tilt tube technique in single determination. 100 μl plasma and 100 μl thromboplastin preparation were mixed and kept at 37 C. After one minute 100 μl 25 mM prewarmed calcium chloride was added and the stop watch started. Tilting was done just above the water line every 1/2 second starting at 8 seconds for normal and 15 seconds for abnormal plasma. The watch was stopped when clotting was detected by visual inspection of the tube.
3.2. VALIDITY OF THE INR STANDARD

3.2.3 Study design

PT was first determined in routine analysis for the patients in November-December 1997, before the plasma was frozen (variable NT\textsubscript{1}). The normal plasmas were not analyzed on this occasion.

Next, on the same day in February 1998, two tubes from each plasma sample were thawed, while the third tube was kept in the freezer for later use. Tubes were successively thawed in 37°C waterbath with frequent but gentle motion to ensure solution of all fibrinogen. When no ice was visible, the tubes were stored at room temperature. Plasma from one tube was used for PT determination by Hepato Quick (variable HQ) and CRM 149S (variable CRM\textsubscript{1}), while plasma from the other tube was used for PT determination by Nycotest (variable NT\textsubscript{2}). The manual tilting was performed by an experienced operator (J.L.H.). The three kinds of measurements were performed independently of each other.

Two months later each sample from the third set of tubes was tested by the manual tilt tube technique using CRM 149S independently by two experienced operators (operator #1 (J.L.H.): variable CRM\textsubscript{2}; operator #2: variable CRM\textsubscript{3}).

3.2.4 Statistical methods

The hypothesis $H_1$ was only tested informally. The slope and intercept were estimated by orthogonal regression for each pair of methods as described in details in the following. In every case it was assumed that the ratio between variances of measurement errors is 1. Next, a plot was made of the estimated regression line with its 95% confidence interval and the individual observed log PTs with their estimated 95% prediction intervals. The plot was examined for outliers and visual deviation from linearity. Moreover, a residual analysis was performed, including probability plots of Best Linear Unbiased Estimators (BLUEs) of residuals to test departure from normality, and plots of BLUEs of residuals versus BLUEs of true x-values to discover departures from linearity and homogeneity of variance.

If $H_1$ was accepted, a formal test was performed for $H_2$. See below for a detailed description of this test. As an important supplement to this test, the mean of the normal log PTs with its 95% confidence ellipse was added to the plot of the abnormal log PTs and the estimated regression line mentioned above. This compact plot provides a good overall visual representation of the calibration exercise, including deviations from any of the two hypotheses in question.

All statistical tests were performed at the 5% level. log always means the
The WHO model for the INR/ISI system

The WHO model states that the relationship between the true log PTs for a pair of methods (thromboplastin × technique) is linear, and that the ratio between the variances of the measurement errors is known. Formally, with the notation of Fuller [117],

\[ y_t = \beta_0 + \beta_1 x_t \]
\[ X_t = x_t + u_t \]
\[ Y_t = y_t + e_t \]
\[ \delta = \frac{\sigma_{ee}}{\sigma_{uu}} \]

for \( t = 1, \ldots, n \) where \( x_t, y_t \) and \( X_t, Y_t \) are the true and observed log PTs, respectively, \( \beta_0 \) and \( \beta_1 \) are the intercept and slope of the line, \( u_t \) and \( e_t \) are the measurement errors, which are assumed to be independent, and \( \sigma_{ee} \) and \( \sigma_{uu} \) are the corresponding variances. This is the classical errors-in-variables model, which dates back to at least the nineteenth century [118, 119].

According to Fuller [117], chapter 1, the method of moments and the method of least squares yield the same estimators of the slope and intercept. The estimates are

\[ \hat{\beta}_1 = \frac{m_{YY} - \delta m_{XX} + [(m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2]^{1/2}}{2m_{XY}} \]
\[ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \]

where

\[ \bar{X} = n^{-1} \sum_{t=1}^{n} X_t \]
\[ \bar{Y} = n^{-1} \sum_{t=1}^{n} Y_t \]
\[ m_{XX} = (n - 1)^{-1} \sum_{t=1}^{n} (X_t - \bar{X})^2 \]
\[ m_{YY} = (n - 1)^{-1} \sum_{t=1}^{n} (Y_t - \bar{Y})^2 \]
\[ m_{XY} = (n - 1)^{-1} \sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y}) \]
3.2. VALIDITY OF THE INR STANDARD

Now, in order to estimate the variances of the estimates, we will assume that \((x_t, e_t, u_t)\) are normally and independently distributed,

\[
(x_t, e_t, u_t) \sim N \left( (\mu_x, 0, 0)', \text{diag}(\sigma_{xx}, \sigma_{ee}, \sigma_{uu}) \right)
\]

Then, according to Fuller, the unbiased method-of-moments estimators of \(\sigma_{xx}, \sigma_{ee}, \text{ and } \sigma_{uu}\) are

\[
\hat{\sigma}_{xx} = (2\delta)^{-1} \left\{ \left[ (m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2 \right]^{1/2} - (m_{YY} - \delta m_{XX}) \right\} \\
\hat{\sigma}_{ee} = \delta \hat{\sigma}_{uu} \\
\hat{\sigma}_{uu} = (2\delta)^{-1} \left\{ m_{YY} + \delta m_{XX} - \left[ (m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2 \right]^{1/2} \right\}
\]

According to Fuller, theorem 1.3.1, an estimate of the asymptotic covariance matrix of the estimators \(\hat{\beta}_0\) and \(\hat{\beta}_1\) is given by

\[
\hat{\text{Var}} \left\{ \hat{\beta}_1 \right\} = (n - 1)^{-1} \hat{\sigma}_{xx}^{-2} \left[ \hat{\sigma}_{xx} s_{vv} + \hat{\sigma}_{uu} s_{vv} - \hat{\sigma}_{uv}^2 \right] \\
\hat{\text{Var}} \left\{ \hat{\beta}_0 \right\} = n^{-1} s_{vv} + \bar{X}^2 \text{Var} \left\{ \hat{\beta}_1 \right\} \\
\hat{\text{Cov}} \left\{ \hat{\beta}_1, \hat{\beta}_0 \right\} = -\bar{X} \text{Var} \left\{ \hat{\beta}_1 \right\}
\]

where

\[
s_{vv} = (n - 2)^{-1} (n - 1) \left( \delta + \hat{\beta}_1^2 \right) \hat{\sigma}_{uu} \\
\hat{\sigma}_{uv} = -\hat{\beta}_1 \hat{\sigma}_{uu}.
\]

The BLUEs of the true \(x\) and the residual, given the observations \(X_t\) and \(Y_t\), are

\[
\hat{x}_t = \left( \hat{\beta}_1^2 \hat{\sigma}_{uu} + \hat{\sigma}_{ee} \right)^{-1} \left( \hat{\sigma}_{uu} (Y_t - \hat{\beta}_0) \hat{\beta}_1 + \hat{\sigma}_{ee} X_t \right) \\
\hat{v}_t = Y_t - \hat{\beta}_0 - X_t \hat{\beta}_1
\]

For model control, \(\hat{v}_t\) and \(\hat{x}_t\) can be treated as the residual and the explanatory variable in an ordinary linear regression. Thus, the usual checks of normality can be performed on \(\hat{v}_t\), and a plot of \(\hat{v}_t\) against \(\hat{x}_t\) can be used to diagnose heteroscedasticity and deviations from the linearity assumption.
A test for the mean normal log PT lying on the line of abnormal plasmas

The following test is based on Tomenson [45].

Assume that we have performed the orthogonal regression in the previous section by using the abnormal plasmas only. Now we wish to test if the mean normal log PT lies on the line of abnormal plasmas.

Let $X_{NT}$ and $Y_{NT}$ denote the observed normal log PTs, $t = 1, \ldots, m$. Let

$$
\bar{X}_N = \frac{1}{m} \sum_{t=1}^{m} X_{NT}
$$

$$
\bar{Y}_N = \frac{1}{m} \sum_{t=1}^{m} Y_{NT}
$$

$$
m_{NXX} = (m - 1)^{-1} \sum_{t=1}^{m} (X_{NT} - \bar{X}_N)^2
$$

$$
m_{NYY} = (m - 1)^{-1} \sum_{t=1}^{m} (Y_{NT} - \bar{Y}_N)^2
$$

$$
m_{NXY} = (m - 1)^{-1} \sum_{t=1}^{m} (X_{NT} - \bar{X}_N)(Y_{NT} - \bar{Y}_N)
$$

A distance from the point $(\bar{X}_N, \bar{Y}_N)$ to the abnormal line is

$$
d = \bar{Y}_N - \hat{\beta}_0 - \hat{\beta}_1 \bar{X}_N
$$

Assume that $(X_{NT}, Y_{NT})'$ are normally and independently distributed with mean $(\mu_{NX}, \mu_{NY})'$. Then the mean of $d$ is

$$
E\{d\} = \mu_{NX} - \beta_0 - \beta_1 \mu_{NY}
$$

which is zero if and only if the point $(\mu_{NX}, \mu_{NY})$ lies on the abnormal line.

The variance of $d$ is approximately (Tomenson p. 108)

$$
\text{Var}\{d\} = \left( m_{NYY} + \beta_1^2 m_{NXX} - 2 \beta_1 m_{NXY} \right) / m
$$

$$
+ \text{Var}\{\beta_0\} + \bar{X}_N^2 \text{Var}\{\beta_1\} + 2 \bar{X}_Nm_{NXY}
$$

For $m$ and $n$ large enough, the variable

$$
Z = d \text{Var}\{d\}^{-1/2}
$$

is approximately standard normally distributed and thus an approximate test value for the hypothesis that the mean normal log PT lies on the abnormal line is $p = 2(1 - \Phi(|Z|))$ where $\Phi$ is the cumulative standard normal distribution.
A confidence ellipse for the mean normal log PT

With the notation and assumptions from the previous sections, let

\[ T^2 = m \left( \bar{X}_N - \mu_{NX} \right) \left( \bar{Y}_N - \mu_{NY} \right) \left( m_{NXX} \quad m_{NXY} \right) \left( \bar{X}_N - \mu_{NX} \right) \left( m_{NXY} \quad m_{NYY} \right) \left( \bar{Y}_N - \mu_{NY} \right) \]

Then (see e.g. Mardia, Kent & Bibby [120], corollary 3.5.1.1) \( T^2 \) has the Hotelling \( T^2 \) distribution with parameters 2 and \( m - 1 \). From theorem 3.5.2 ibid. it follows that

\[ T^2(2, m - 1) = \frac{2(m - 1)}{m - 2} F_{2, m - 1} \]

where \( F_{2, m - 1} \) is the \( F \)-distribution with 2 and \( m - 1 \) degrees of freedom.

Now, a \((1 - \alpha)\) confidence ellipse for \((\mu_{NX}, \mu_{NY})'\) is given by

\[ \left\{ \left( \begin{array}{c} \mu_{NX} \\ \mu_{NY} \end{array} \right) \mid T^2 \leq t^2 \right\} \]

where \( t^2 \) is the \((1 - \alpha)\) fractile in the \( T^2(2, m - 1) \) distribution. Solving the inequality yields the solutions

\[ C - D^{1/2} \leq \mu_{NY} \leq C + D^{1/2} \]

\[ \bar{X}_N - \left( \frac{t^2}{m} m_{NXX} \right)^{1/2} \leq \mu_{NX} \leq \bar{X}_N + \left( \frac{t^2}{m} m_{NXX} \right)^{1/2} \]

where

\[ C = \bar{Y}_N + \frac{m_{NXY}}{m_{NXX}} (\mu_{NX} - \bar{X}_N) \]

\[ D = \frac{m_{NXY} - m_{NXX} m_{NYY}}{m_{NXX}} \left( \frac{(\mu_{NX} - \bar{X}_N)^2}{m_{NXX}} - \frac{t^2}{m} \right) \]
3. MATERIALS AND METHODS OF EMPIRICAL STUDIES
4

Results of empirical studies

4.1 Precision of home coagulometer

No patient experienced major bleedings or embolic events. Each of the 13 patients reported INR values from 10 or 11 double measurements on his/her coagulometer using lot 44 of test carriers. The median interval between double measurements was 7 days (range 5-9 days). Moreover, simultaneous double determinations using lots 44 and 45 were obtained from 12 patients; one patient only performed a single measurement on each lot. Based on average home coagulometer measurements, patients had median INR levels ranging from 2.15 to 3.25 (figure 4.1), and 100 out of 153 INR values (65%) were within the therapeutic target range. There were no obvious outlier data, and all observed results were included in the statistical calculations.

The memory was copied from all coagulometers except for one, which was empty due to missing batteries. Due to extra control measurements, one coagulometer contained only 19 out of 22 reported INR values. One patient had discordance between two results, both on the same day (patient entered 2.2 and 2.3, whereas coagulometer memory displayed 2.3 and 2.1, respectively). All other measurements were concordant. In total there was agreement in 253 out of 255 INR values (99.2%).

4.1.1 Precision

Out of 152 double measurements of INR performed by patients using lot 44 the differences in 53 were within $\pm 0.1$ INR, 138 were within $\pm 0.3$ INR and all were within $\pm 0.7$ INR. In figure 4.2 (left) the differences between first and second measurements are plotted against the averages of the two measurements, demonstrating increasing variance with increasing levels of INR. Figure 4.2 (right) shows how the variance is stabilized by the use of relative
The mean relative difference between first and second measurement was 0.09% (CI -1.1% to 1.3%). The CV of single measurements was 5.2% (CI 4.7% to 5.9%), and there was no significant difference in precision between patients ($p = 0.27$).

Out of 12 double determinations of INR performed by patients using lot 45, the differences between first and second measurement in 9 were within ±0.1 INR, two were -0.2 INR and one was -0.4 INR.

Out of 12 double measurements on the two lots (average of 2 measurements) 7 were within ±0.1 INR and all were within ±0.25 INR except one, where the difference between lot 45 and lot 44 was 0.45 INR. The mean relative difference between lot 45 and lot 44 was 3.2% (CI -0.2% to 6.5%). Considering only simultaneous measurements using lots 44 and 45 the CV of single measurements was 6.1% (CI 5.0% to 7.9%).

Considering all measurements on both lots (corrected for different patient levels and serial correlations), the CV of single measurements was 5.4% (CI 4.9% to 6.1%).
4.1. PRECISION OF HOME COAGULOMETER

4.1.2 Comparison of home coagulometers with routine hospital laboratories

The first series of comparisons in June 1997 showed a mean relative difference of 24.3% (CI 17.4% to 31.2%) between Skejby Sygehus and home coagulometer INRs with individual differences (average of two measurements) ranging from 0.25 INR to 1.25 INR (figure 4.3). The second series of comparisons in August 1997 revealed mean relative differences of 17.1% (CI 11.1% to 23.0%) between Skejby Sygehus and home coagulometers, 6.7% (CI 0.8% to 12.7%) between Roskilde Amtssygehus and home coagulometers, and 10.4% (CI 4.7% to 16.0%) between Skejby Sygehus and Roskilde Amtssygehus (figure 4.4).

The variance of the estimate of the relative difference between Skejby Sygehus and CoaguChek has four components: the long-term between-day analytical CV of each method (7.1% and 8.4%), and the uncertainty of the day-specific estimate due to within-day analytical variation within each method (3.6% and 5.4%). The numbers in parentheses are estimates for Skejby Sygehus and CoaguChek respectively, based on results from the present and an earlier study [116]. Applying these estimates, the CV of the estimate of the relative difference due to long-term between-day analytical variation
is $(0.0712^2 + 0.0842^2)^{1/2} = 11.0\%$. Likewise, the CV of the estimate due to within-day analytical variation is \(\{(0.0362^2 + 0.0542^2)/(2 \times 13)\}^{1/2} = 1.3\%\). Thus, the total long-term CV of the estimate is \((0.1102^2 + 0.0132^2)^{1/2} = 11.1\%\). In our earlier study we estimated that the long-term relative difference between Skejby Sygehus and CoaguChek was 8\%. Assuming that this is the true value, the probability of observing a relative difference of 24.3\% or worse is 14\% (two-tailed test).

### 4.2 Validity of INR standard

#### 4.2.1 Marginal results

The geometric means (GM) and coefficients of variation (CV) of the GM of the PTs obtained by the different thromboplastins are shown in table 4.1.

The measurements by CRM 149S were analyzed by analysis of variance with sample and sample \(\times\) period as random effects. Period denotes the time of measurement, with CRM\(_1\) measured in period 1, and CRM\(_2\) and
CRM3 measured in period 2. For normal plasmas the PT was prolonged with 1.5% (se 1.8%) from first to second period. The prolongation was, however, not statistically significant \((p = 0.42)\). For abnormal plasmas the PT was significantly prolonged \((p < 0.0001)\) with 3.5% (se 0.5%).

### 4.2.2 The linearity hypotheses

Figure 4.5 is a scatterplot matrix of all the measurements performed in the study. The plot gives an overall view of the relations between the log PTs obtained by each pair of methods.

Figures 4.6-4.7 are compact plots revealing the most important aspects of the data for each pair of methods. log PTs in log(seconds) of abnormal plasmas are plotted as individual data points. The mean normal log PT is shown as a circle within its 95% confidence ellipse. The solid line is the estimated orthogonal regression line, based on abnormal plasmas only. The punctuated lines represent the 95% confidence interval for the orthogonal regression line, while the dotted lines are the 95% prediction interval for abnormal log PT. See the legend in figure 4.7.
Table 4.1: Geometric means (GM) and coefficients of variation of geometric means (CV(GM)).

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th></th>
<th>Abnormal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GM seconds</td>
<td>CV(GM) %</td>
<td>GM seconds</td>
<td>CV(GM) %</td>
</tr>
<tr>
<td>NT₁</td>
<td>50.77</td>
<td>3.74</td>
<td>50.77</td>
<td>3.74</td>
</tr>
<tr>
<td>NT₂</td>
<td>20.72</td>
<td>2.40</td>
<td>44.47</td>
<td>3.98</td>
</tr>
<tr>
<td>HQ</td>
<td>22.16</td>
<td>2.40</td>
<td>50.76</td>
<td>3.93</td>
</tr>
<tr>
<td>CRM₁</td>
<td>16.07</td>
<td>1.94</td>
<td>28.79</td>
<td>3.01</td>
</tr>
<tr>
<td>CRM₂</td>
<td>16.32</td>
<td>2.50</td>
<td>29.55</td>
<td>3.33</td>
</tr>
<tr>
<td>CRM₃</td>
<td>16.31</td>
<td>2.62</td>
<td>29.32</td>
<td>3.32</td>
</tr>
</tbody>
</table>

For all pairs of methods, the log PTs of abnormal plasmas had no outliers or visual deviations from linearity, except for two observations of CRM₁ against CRM₂, and the same of CRM₁ against CRM₃ (see figure 4.6). With this exception, the examination of BLUEs of residuals showed no deviations from the assumption of normality. With the same exception, plots against BLUEs of the true x-values showed no deviations from the assumption of homogeneity of variances nor from linearity. Thus, data were found to be consistent with \( H₁ \) for all pairs of methods, except for CRM₁ against CRM₂, and for CRM₁ against CRM₃. It is, however, surprising that a group of size 4-7 of abnormal plasmas lies in or close to the normal range, clearly separated from the remaining abnormal plasmas. This will be further discussed later.

In order to examine the influence of the two outliers in the two rejected pairs of methods, new calculations for these two pairs were performed without the outliers. For CRM₁ versus CRM₂, the slope estimate changed from 0.961 ± 0.027 to 0.944 ± 0.019, and for CRM₁ versus CRM₃ the slope estimate changed from 0.962 ± 0.022 to 0.948 ± 0.019.

Except for the Nycotest measurements in November–December 1997, where no normal plasmas were measured, the formal tests showed that data were consistent with \( H₂ \) for each pair of methods with p-values ranging from 0.08 to 0.99 (0.08 to 0.72 without outliers). This result is supported by the compact plots.

The results are summarized in table 4.2.
4.2. VALIDITY OF INR STANDARD

Figure 4.5: Scatterplot matrix of PT determinations in log(seconds) for Nycotest in December 1997, Hepato Quick, Nycotest and CRM 149S in February 1998, and CRM149S in April 1998. The measurements in April 1998 were performed by two different operators.

4.2.3 INR values

The Nycotest system at the Department of Clinical Biochemistry, Skejby Sygehus, February 1998

The average slope of the three orthogonal regression lines relating CRM 149S to the local Nycotest system of February 1998 is 0.818 (see table 4.2). Assuming that the three sets of measurements by CRM 149S are independent, the se of the average slope is 0.016. CRM 149S was assigned an ISI of 1.257 with se 0.013 [18]. Thus an estimate of the ISI of Nycotest, February 1998 is $0.818 \times 1.257 = 1.028$. The stated SE of CRM 149S is probably too low (see chapter 5), but it allows us to calculate a lower bound of the SE of the ISI of Nycotest. The se of the ISI estimate for Nycotest is, therefore $\geq (0.818^2 \times 0.013^2 + 0.016^2 \times 1.257^2 + 0.016^2 \times 0.013^2)^{1/2} = 0.023$. Thus, compared with the ISI of 1.04 assigned to the local Nycotest system by the
Figure 4.6: Plots from calibrations. See the text for details.
4.2. VALIDITY OF INR STANDARD

Figure 4.7: Plots from calibrations. See the text for details.
Table 4.2: Estimates of slopes with standard errors and p-values for testing the hypothesis $H_2$. Numbers in parentheses are estimates and p-values without two outlying observations (see text for details).

<table>
<thead>
<tr>
<th></th>
<th>CRM$_1$</th>
<th>CRM$_2$</th>
<th>CRM$_3$</th>
<th>HQ</th>
<th>NT$_2$</th>
<th>NT$_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>SE</td>
<td>SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.961</td>
<td>0.027</td>
<td>0.022</td>
<td>0.754</td>
<td>0.031</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>0.944</td>
<td>0.019</td>
<td>0.019</td>
<td>0.948</td>
<td>0.031</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.47</td>
<td>0.78</td>
<td>0.40</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>1.004</td>
<td>0.015</td>
<td>0.028</td>
<td>0.787</td>
<td>0.025</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>0.13</td>
<td>0.13</td>
<td>0.787</td>
<td>0.12</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>0.786</td>
<td>0.029</td>
<td>0.025</td>
<td>0.828</td>
<td>0.032</td>
<td>0.783</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.952</td>
<td>0.009</td>
<td>1.000</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.945</td>
<td>0.016</td>
<td>0.945</td>
<td>0.016</td>
<td>0.945</td>
</tr>
</tbody>
</table>

By insertion it is found that the mean relative difference in INR between Nycotest and CRM 149S is 4–5% for INR in the range 2–4.
4.2. VALIDITY OF INR STANDARD

The local Hepato Quick system, February 1998

The ISI calculated from the three orthogonal regression lines of CRM 149S against Hepato Quick is 0.975 with se 0.021. Thus, there is no statistically significant difference between this and the assigned ISI value of 0.99. The MNPT based on the 10 normal samples was 22.16 s with se 0.53 s.

In the same way as above it is found that the mean relative difference between INR measured by Hepato Quick and CRM 149S is 6-7%.
4. RESULTS OF EMPIRICAL STUDIES
Accuracy of INR estimates: mathematical derivation

5.1 A motivating example: The calibration of the International Reference Preparation RBT/79

In 1983, Hermans et al. [13] published a report on the calibration of the thromboplastin preparation RBT/79 (rabbit, plain) against IRP 67/40 (human, combined). Five thromboplastins, including RBT/79, were calibrated against the WHO primary IRP 67/40 in a collaborative study of ten European and American laboratories. The calibration was performed according to WHO recommendations and the slope of the regression line of IRP 67/40 on RBT/79 was estimated by orthogonal regression (see also page 25) in each laboratory. RBT/79 was then assigned an ISI value equal to the average of the estimates from the seven European laboratories, ISI = 1.413.

In figure 5.1 the dashed line represents the average of the slopes estimated by orthogonal regression lines of the seven laboratories with slope equal to the assigned ISI 1.413. Now suppose, for example, that for a given patient the prothrombin time (PT) measured in the first laboratory is 35 seconds. The mean normal prothrombin time (MNPT) in laboratory 1 estimated from 20 normal plasmas was reported to be 14.63 seconds and applying the assigned ISI the estimated INR would then be calculated as \( \frac{35}{14.63}^{1.413} = 3.43 \).

The idea of using the average of the slopes is to borrow strength across laboratories to increase the precision of the common slope estimate. This is, however, only feasible if all the laboratories share the same true slope. Consider a situation where the true slopes are unequal. Since INR is defined
as the PT ratio that would have been obtained, had the IRP 67/40 been used for PT determinations, the best unbiased estimator in this case would be obtained by using the slope of laboratory no. 1. This is also depicted in figure 5.1 where the solid line represents the estimated orthogonal regression line of laboratory no. 1 with slope 1.482. By using the laboratory specific slope, the estimate of the INR would be \((35/14.63)^{1.482} \approx 3.64\).

Using a laboratory specific slope will of course, other things being equal, lead to a lower precision of the INR estimate, but this only reflects the reality and is the price to be paid in order to obtain an unbiased estimate. If the variation between laboratories is not considered, the estimates of INR may be assigned an unrealistic high precision. This can seriously damage the INR standard and may lead to conflicting results in international studies and to lack of a firm basis for dosage recommendations in oral anticoagulant therapy.

In the study by Hermans et al, there is considerable evidence that the slopes in fact are unequal. This will be further discussed in section 5.5.1.
In the following section we will propose a statistical model, which incorporates the possibility of unequal slopes across laboratories. The model can be used to evaluate the accuracy of INR estimates originating from various IRPs.

5.2 A contemporary model for INR

In contemporary calibration methodology, the model behind the ISI determinations is the classical errors-in-variables model. When the variances of the measurement errors on both axes are equal (as is commonly assumed in ISI calibrations), the procedure for estimating the parameters in this model is called orthogonal regression (see page 48 ff).

The slope estimates are calculated by orthogonal regression by using all the data points without distinguishing between normal and abnormal plasmas. However, in the present substudy, we have not considered the methodology behind the calculation of slope estimates and their precision, but in order to derive distributional results for the interlaboratory variation we have only assumed that the slope estimates and their variance estimates are independent and that the distribution of the variance estimate is approximately a scaled $\chi^2$.

5.2.1 A semi-deterministic model

In this section we will present a formal description of the model, which underlies the WHO recommendation [102]. Let $\mathcal{P}$ and $\mathcal{S}$ be sets of patients and normal persons, respectively. Let $\mathcal{T}$ denote a set of thromboplastin preparations (with meticulously specified measurement methods) and assume that for any pair $(t_1, t_2) \in \mathcal{T}^2$ there exist real numbers $a_{12}$ and $b_{12}$ such that for any patient $p \in \mathcal{P}$

$$x_{1p} = a_{12} + b_{12}x_{2p} \quad (5.1)$$

where $x_{ip}$ is the “true” logarithmic prothrombin time obtained by $t_i$, $i = 1, 2$. Assume that for a randomly selected normal person $s \in \mathcal{S}$, the “true” log(PT) measured by $t_i$, $i = 1, 2$, is

$$z_{is} \sim N \left( \mu_i, \tau_i^2 \right)$$

and that the point corresponding to the means of the logarithmic prothrombin times for normal persons lies on the line given by (5.1), i.e.

$$\mu_1 = a_{12} + b_{12}\mu_2$$
Then, for any patient \( p \),
\[
 x_{1p} - \mu_1 = b_{12} (x_{2p} - \mu_2)
\]

Thus, if the true slope \( b_{12} \) is known, the thromboplastin \( t_2 \) can be used to find the difference between the patient’s and the mean normal logarithmic prothrombin times that would have been obtained by using the thromboplastin \( t_1 \).

For a given patient \( p \), the International Normalized Ratio (INR) is defined as the exponential function of the difference \( x_{0p} - \mu_0 \) that would have been obtained, had the thromboplastin \( t_0 \) been the 67/40. Let \( y_p \) denote the true log(INR) for patient \( p \) and assume that IRP 67/40 with its specified method belongs to \( T \). Then for any thromboplastin \( t_i \in T \) there exists a real number \( c_i \) such that
\[
y_p = x_{0p} - \mu_0 = c_i (x_{ip} - \mu_i)
\]

The constant \( c_i \) is called the International Sensitivity Index (ISI) of the thromboplastin \( t_i \).

Now let \( t_0, \ldots, t_r \in T \) be a sequence of thromboplastins, with \( b_{i-1,i} \) being the slope of the regression line of the logarithmic prothrombin times of \( t_{i-1} \) on \( t_i \). Assume that \( t_0 \) is the IRP 67/40. Under the conditions given above, the ISI of \( t_r \) is
\[
c_r = \prod_{i=1}^{r} b_{i-1,i}
\]
and thus the INR can be calculated from prothrombin times obtained by the thromboplastin \( t_r \),
\[
y_p = c_r (x_{rp} - \mu_r)
\]

### 5.2.2 Extension to a statistical model

When uncertainty is introduced, the ISI estimate will depend on the path of calibration. In the following we will consider only one path \( t_0, \ldots, t_r \in T \) of calibration and instead of writing \( "i-1,i" \) in the subscripts we will simply write \( "i" \). Similarly we will suppress \( "r" \) in the notation.

Let \( L \) denote a set of laboratories, and assume that the true slope in the \( l \)'th laboratory is
\[
b_{il} = b_i + \gamma_i U_{il}, \quad i = 1, \ldots, r, \quad l \in L
\]
where \( b_i \) is the common true slope, \( \gamma_i^2 \) is the inter-laboratory variance of the true laboratory-specific slopes and \( U_{il} \) is standard normally distributed. We
allow the true slopes $b_{il}$ and $b_{i'l'}$ within each laboratory $l$ to be correlated with correlation coefficient $\rho_{ii'l'}$, but assume independence across laboratories, i.e.

$$\text{corr}(b_{il}, b_{i'l'}) = \text{corr}(U_{il}, U_{i'l'}) = \begin{cases} \rho_{ii'l'}, & l = l' , \ i, i' = 1, \ldots, r, \ l, l' \in \mathcal{L} \\ 0, & l \neq l' \end{cases}$$

where $\{\rho_{ii'l'}\}$ must be chosen such that the covariance matrix of $\{U_{il}\}$ is positive definite.

Suppose for each thromboplastin $t_i$ that each of the $k_i$ laboratories in a set $L_i \subseteq \mathcal{L}$ estimates the common true slope $b_i$ of the regression line of $t_{i-1}$ on $t_i$ by

$$\hat{b}_{il} = b_{il} + \nu_{il}V_{il}, \ i = 1, \ldots, r, \ l \in L_i$$

where $\nu_{il}^2$ is the variance of the estimation error in the $l$'th laboratory, given the mean slope $b_{il}$, and $V_{il}$ is standard normally distributed. Then the commonly used unweighted estimate of the slope $b_i$ is

$$\tilde{b}_i = \frac{1}{k_i} \sum_{l \in L_i} \hat{b}_{il}, \ i = 1, \ldots, r \quad (5.3)$$

The ISI of $t_r$ is

$$c = \prod_{i=1}^{r} b_i$$

and the common estimator of $c$ is

$$\tilde{c} = \prod_{i=1}^{r} \tilde{b}_i$$

Now, let $\lambda$ denote a randomly chosen laboratory, which did not participate in any calibration exercise, i.e.

$$\lambda \in \mathcal{L} \setminus \left( \bigcup_{i=1}^{r} L_i \right) \quad (5.4)$$

In the following we assume that the fixed laboratory $\lambda$ uses the thromboplastin $t_r$ to perform all PT determinations. The mean normal PT is estimated from a random selection of $n$ normal persons in $S$. Suppose that the mean log(PT) of the normal population is $\mu$ and that the true log(PT) of the $s$'th normal person is

$$z_s = \mu + \tau W_s, \ s = 1, \ldots, n$$
where $\tau^2$ is the variance between normal persons and $W_s$ is standard normally distributed. For each normal person, a single PT determination is performed, and the observed log(PT) is

$$\tilde{z}_s = z_s + \sigma \delta_s, \ s = 1, \ldots, n$$

where $\sigma^2$ is the variance of the analytical error in PT determinations and $\delta_s$ is standard normal. The mean normal log(PT) is estimated by

$$\tilde{\mu} = n^{-1} \sum_{s=1}^{n} \tilde{z}_s \quad (5.5)$$

For a given patient assume that the true log(PT) is $x$ and that the observed log(PT) is

$$\tilde{x} = x + \sigma \varepsilon \quad (5.6)$$

where $\varepsilon$ is standard normal. Let $y$ denote the true log(INR) and define, in accordance with (5.2), the $t_r$-estimator of $y$ as

$$\tilde{y} := c (\tilde{x} - \tilde{\mu}) \quad (5.7)$$

For the errors we assume that

$$\left\{ \left\{ (U_{il}, \ldots, U_{rl})' \right\}_l, \{V_{il}\}_l, \{W_s\}_s, \{\delta_s\}_s, \varepsilon \right\} \quad (5.8)$$

are independent.

Note that the true log(INR) is

$$y = c_\lambda (x - \mu)$$

where

$$c_\lambda = \prod_{i=1}^{r} b_{i,\lambda} = \prod_{i=1}^{r} b_i \prod_{i=1}^{r} \left( 1 + \frac{\gamma_i}{b_i} U_{i,\lambda} \right) = c \prod_{i=1}^{r} \left( 1 + \frac{\gamma_i}{b_i} U_{i,\lambda} \right) \quad (5.9)$$

is the true ISI of the laboratory $\lambda$.

In the next section we will find the distribution of the estimated log(INR), $\tilde{y}$, given the true log(INR), $y$, when the variances $\gamma_i^2$, $\nu_{il}^2$, and $\sigma^2$ tend to zero, and the number of normal persons, $n$, tends to infinity.

5.3 Distribution of the INR estimator

**Theorem 1** Assume that the model in section 5.2.2 holds and that there exist positive real numbers $\dot{\gamma}_i$, $\dot{\nu}_{il}$, $\dot{\sigma}$, and $M$, $i = 1, \ldots, r$, $l \in L$, such that for every $h > 0$, $\gamma_i = h \dot{\gamma}_i$, $\nu_{il} = h \dot{\nu}_{il}$, $\sigma = h \dot{\sigma}$, and $n^{-1/2} = h M$. 


5.3. DISTRIBUTION OF THE INR ESTIMATOR

Define

\[ \zeta_r^2 = \sum_{i=1}^{r} \sum_{i' = 1}^{r} \frac{\gamma_i \gamma_{i'}}{b_i k_i b_{i'} k_{i'}} \sum_{l=l'}^{\gamma_r - 1} \rho_{i'i} + \sum_{i=1}^{r} \frac{1}{b_i^2 k_i} \nu_i^2 \]  

(5.10)

\[ = \zeta_r^2 - 1 + \gamma_r^2 + \nu_r^2 + 2 \sum_{i=1}^{r-1} \frac{\gamma_i \gamma_r}{b_i k_i b_r k_r} \sum_{l=l'}^{\gamma_r - 1} \rho_{i'l} \]

(5.11)

\[ \kappa_r^2 = \sum_{i=1}^{r} \sum_{i' = 1}^{r} \frac{\gamma_i \gamma_{i'}}{b_i b_{i'}} \rho_{i'\lambda} \]

(5.12)

\[ \omega^2 = c^2 \left( \sigma^2 + \frac{\tau^2}{n} \right) \]

(5.13)

where \( \nu_i^2 = 1/k_i \sum_{l \in L_i} \nu_{il}^2 \) is the average variance of the slope estimates in the \( i \)'th calibration, and \( \kappa_0^2 = \zeta_0^2 = 0 \).

Then, for fixed true \( \log(\text{INR}) = y > 0 \), \( b_i > 0 \), \( i = 1, \ldots, r \), and conditional on \( \{c_\lambda > 0, \hat{x} > \hat{\mu}, b_i > 0, i = 1, \ldots, r\} \), we have the following results on convergence in law:

\[ \frac{\log (\hat{c}) - \log (c)}{\zeta_r} \xrightarrow{\text{d}} N(0,1) \quad \text{for } h \downarrow 0 \]  

(5.14)

\[ \frac{\log (c_\lambda) - \log (c)}{\kappa_r} \xrightarrow{\text{d}} N(0,1) \quad \text{for } h \downarrow 0 \]  

(5.15)

\[ \frac{\log (\hat{x} - \hat{\mu}) - \log (y/c)}{(\kappa_r^2 + \omega^2/y^2)^{1/2}} \xrightarrow{\text{d}} N(0,1) \quad \text{for } h \downarrow 0 \]  

(5.16)

\[ \frac{\log (\hat{y}) - \log (y)}{\hat{y}} \xrightarrow{\text{d}} N(0,1) \quad \text{for } h \downarrow 0 \]  

(5.17)

Remark:

- The results in the theorem should not be used to estimate the relative bias of INR estimates, since this is of the same magnitude as the model uncertainty. See section 5.3.1 for details.
- The theorem shows that the inaccuracy of estimates of INR is a function of
– \( \zeta_r \): approximately the CV of the common ISI estimate, \( \hat{c} \),
– \( \kappa_r \): approximately the CV of the true ISI between laboratories, \( c_\lambda \),
– \( \tau \): approximately the CV of true PTs between normal persons, and
– \( \sigma \): approximately the \( \lambda \)-specific CV of the short-term analytical error of PT (the intra-laboratory CV).

\( \zeta_r \) can be reduced by increasing the number of laboratories participating in the calibrations. The degree, to which it can be reduced, depends, however, on the correlation between calibrations within each laboratory. Another way to reduce \( \zeta_r \) is to choose different laboratories for different calibrations.

\( \kappa_r \) is due to the fact that each laboratory in reality has its own specific true ISI value (as demonstrated in figure 5.1). It cannot be reduced by increasing the number of participating laboratories and/or the number of PT determinations; the only way to reduce \( \kappa_r \) is to strictly harmonize the method of PT determinations by IRPs.

\( \omega \) can be reduced by increasing the number of normals used for the \( \lambda \)-specific estimation of the mean normal PT, and by specifying stricter selection criteria for normal persons.

Proof:

\( (5.14) \): Write

\[
\log \left( \frac{\hat{c}}{c} \right) = \log \left( \prod_{i=1}^{r} \frac{b_i}{\hat{b}_i} \right)
= \sum_{i=1}^{r} \log \left( \frac{\hat{b}_i}{b_i} \right)
= \sum_{i=1}^{r} \log \left( k_i^{-1} b_i^{-1} \sum_{l \in L_i} (b_i + \gamma_i U_{il} + \nu_i V_{il}) \right)
= \sum_{i=1}^{r} \log \left( \frac{\gamma_i}{b_i k_i} \sum_{l \in L_i} U_{il} + \frac{1}{b_i k_i} \sum_{l \in L_i} \nu_i V_{il} \right)
= \sum_{i=1}^{r} \log \left( \frac{\hat{\gamma}_i}{b_i k_i} \sum_{l \in L_i} U_{il} + \frac{1}{b_i k_i} \sum_{l \in L_i} \hat{\nu}_i V_{il} \right)
\]

\( (5.18) \)
The term (5.18) tends to zero, almost surely, as } h \downarrow 0 \text{ and is differentiable with respect to } h \text{ with derivative }

\frac{\partial \{ \log (\tilde{c}/c) \}}{\partial h} = \sum_{i=1}^{r} \frac{\dot{\gamma}_i}{b_i k_i} \sum_{l \in L_i} U_{il} + \frac{1}{b_i k_i} \sum_{l \in L_i} \dot{v}_l V_{il}

\rightarrow \sum_{i=1}^{r} \frac{\dot{\gamma}_i}{b_i k_i} \sum_{l \in L_i} U_{il} + \sum_{i=1}^{r} \frac{1}{b_i k_i} \sum_{l \in L_i} \dot{v}_l V_{il} \quad \text{a.s.} \tag{5.19}

for } h \downarrow 0 \text{. The distribution of this limit is normal with mean 0 and variance } \dot{\zeta}^2_r = \zeta^2_r / h^2 \text{. Moreover, since } \zeta_r \to 0 \text{ for } h \downarrow 0, \text{ and } \zeta_r \text{ is differentiable with respect to } h \text{ with derivative } \zeta_r \text{ it follows from l'Hospital's rule that}

\lim_{h \downarrow 0} \frac{\log (\tilde{c}/c)}{\zeta_r} = \zeta_r^{-1} \left\{ \sum_{i=1}^{r} \frac{\dot{\gamma}_i}{b_i k_i} \sum_{l \in L_i} U_{il} + \sum_{i=1}^{r} \frac{1}{b_i k_i} \sum_{l \in L_i} \dot{v}_l V_{il} \right\} \quad \text{a.s.}

Now, the desired result follows from the fact that almost sure convergence implies convergence in law.

(5.15): By (5.9) we can write

\log \left( c_{\lambda}/c \right) = \log \left\{ \prod_{i=1}^{r} \left( 1 + \frac{\gamma_i}{b_i} U_{i\lambda} \right) \right\}

= \sum_{i=1}^{r} \log \left( 1 + \frac{\gamma_i}{b_i} U_{i\lambda} \right)

= \sum_{i=1}^{r} \log \left( 1 + \frac{h \gamma_i}{b_i} U_{i\lambda} \right)

This expression tends to 0 as } h \downarrow 0 \text{, a.s., and is differentiable with respect to } h \text{ with derivative }

\frac{\partial \{ \log (c_{\lambda}/c) \}}{\partial h} = \sum_{i=1}^{r} \frac{\dot{\gamma}_i / b_i}{1 + h \frac{\dot{\gamma}_i}{b_i} U_{i\lambda}}

\rightarrow \sum_{i=1}^{r} \frac{\dot{\gamma}_i}{b_i} U_{i\lambda} \quad \text{a.s. for } h \downarrow 0

The distribution of this limit is normal with mean 0 and variance } \kappa^2_r = \kappa^2_r / h^2 \text{. Thus, by l'Hospital's rule,

\lim_{h \downarrow 0} \frac{\log (c_{\lambda}/c)}{\kappa_r} = \kappa_r^{-1} \sum_{i=1}^{r} \frac{\dot{\gamma}_i}{b_i} U_{i\lambda} \quad \text{a.s.}
ACCURACY OF INR ESTIMATES

(5.16): Note that
\[ \frac{c}{y} (\tilde{x} - \tilde{\mu}) \]
\[ = \frac{c}{y} (x + \sigma \varepsilon - \tilde{\mu}) \]
\[ = \frac{c}{c_\lambda} (x - \tilde{\mu}) \frac{c}{c_\lambda} + \frac{c}{y} (\sigma \varepsilon - \tilde{\mu} + \mu) \]
\[ = \frac{c}{c_\lambda} + \frac{c}{y} (\sigma \varepsilon - \tilde{\mu} + \mu) \]
\[ = \prod_{i=1}^{r} \left( 1 + \frac{\gamma_i}{b_i} U_{i\lambda} \right)^{-1} \frac{c}{y} \left\{ \sigma \varepsilon - n^{-1} \sum_{s=1}^{n} (\mu + \tau W_s + \sigma \delta_s) + \mu \right\} \]
\[ = \sum_{i=1}^{r} \frac{\gamma_i}{b_i} U_{i\lambda} \left( 1 + n^{-1/2} \frac{\gamma_i}{M b_i} U_{i\lambda} \right)^{-1} \frac{c}{y} \left\{ \sigma \varepsilon - n^{-1} \sum_{s=1}^{n} \delta_s - \tau n^{-1} \sum_{s=1}^{n} W_s \right\} \]
\[ = Z_{1n} + Z_{2n} \quad (5.20) \]

where, by assumption, \( n = n(h) = \frac{h^{-2} M^2}{C} \).

First,
\[ \frac{\partial Z_{1n}}{\partial n} = \frac{1}{2} n^{-3/2} Z_{1n} \sum_{i=1}^{r} \frac{\gamma_i}{M b_i} U_{i\lambda} \left( 1 + n^{-1/2} \frac{\gamma_i}{M b_i} U_{i\lambda} \right)^{-1} \]

Thus, by l'Hospital's rule,
\[ \lim_{n \to \infty} \frac{Z_{1n} - 1}{n^{-1/2}} = \lim_{n \to \infty} \frac{\partial \{ Z_{1n} - 1 \} / \partial n}{\partial \{ n^{-1/2} \} / \partial n} \]
\[ = \lim_{n \to \infty} -Z_{1n} \sum_{i=1}^{r} \frac{\gamma_i}{M b_i} U_{i\lambda} \left( 1 + n^{-1/2} \frac{\gamma_i}{M b_i} U_{i\lambda} \right)^{-1} \]
\[ = - \sum_{i=1}^{r} \frac{\gamma_i}{M b_i} U_{i\lambda} \quad \text{a.s.} \]

Therefore,
\[ n^{1/2} (Z_{1n} - 1) \approx N \left( 0, \kappa_{\sigma^2}^2 / M^2 \right) \]

Next, let \( \dot{\omega} = \omega / h = c (\dot{\sigma}^2 + \tau^2 M^2)^{1/2} \). Then
\[ \lim_{n \to \infty} n^{1/2} Z_{2n} \]
\[ = \lim_{n \to \infty} n^{1/2} \frac{c}{y} \left\{ M^{-1} n^{-1/2} \dot{\sigma} \left( \varepsilon - n^{-1} \sum_{s=1}^{n} \delta_s \right) - \tau n^{-1} \sum_{s=1}^{n} W_s \right\} \]
\[ = \lim_{n \to \infty} \frac{c}{y} \left\{ M^{-1} \dot{\sigma} \left( \varepsilon - n^{-1} \sum_{i=1}^{n} \delta_s \right) - \tau n^{-1/2} \sum_{s=1}^{n} W_s \right\} \]
\[ = \frac{c}{y} \left\{ \dot{\sigma} / M \varepsilon - \tau \lim_{n \to \infty} n^{-1/2} \sum_{s=1}^{n} W_s \right\} \quad \text{a.s.} \]
5.3. DISTRIBUTION OF THE INR ESTIMATOR

Therefore,
\[ n^{1/2}Z_{2n} \xrightarrow{\mathcal{D}} N\left(0, \frac{\omega^2}{y^2M^2}\right) \text{ for } n \to \infty \]

Consider the function \((z_1, z_2) \mapsto \log (z_1 + z_2)\) for \(z_1 + z_2 > 0\). By a Taylor expansion about \((1, 0)\), we get
\[ \log (z_1 + z_2) = z_1 - 1 + z_2 + \psi (z_1, z_2) \left\| (z_1 - 1, z_2) \right\| \]
where \(\psi (z_1, z_2) \to 0\) for \((z_1, z_2) \to (1, 0)\).

Now, by insertion, we find that
\[ n^{1/2} \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\} = n^{1/2} \log (Z_{1n} + Z_{2n}) \xrightarrow{\mathcal{D}} N \left(0, \left(\kappa_r^2 + \omega^2/y^2\right)/M^2\right) \text{ for } n \to \infty \]

Finally,
\[ \frac{\log (\hat{x} - \hat{\mu}) - \log (y/c)}{\left(\kappa_r^2 + \omega^2/y^2\right)^{1/2}} = \frac{n^{1/2} \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\}}{\left\{ \kappa_r^2 + \omega^2/y^2 \right\}^{1/2}/M} \xrightarrow{\mathcal{D}} N(0, 1) \text{ for } h \downarrow 0 \]

(5.17): Write
\[ \log (\hat{y}) - \log (y) = \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\} - \log (y) = \log (\hat{c}/c) + \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\} \]

Note that the terms are independent due to the assumptions (5.4) and (5.8).
Let \(\hat{\theta} = \theta/h\). Then, by (5.21), (5.14), and (5.16),
\[ \frac{\log (\hat{y}) - \log (y)}{\hat{\theta}} = \frac{\log (\hat{c}/c) + \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\}}{h\hat{\theta}} \]
\[ = \hat{\theta}^{-1} \left\{ \frac{\hat{\zeta}_r \log (\hat{c}/c)}{\zeta_r} + \left(\kappa_r^2 + \omega^2/y^2\right)^{1/2} \log \left\{ \frac{c}{y} (\hat{x} - \hat{\mu}) \right\} \right\} \]
\[ \quad \xrightarrow{\mathcal{D}} N \left(0, \frac{\hat{\zeta}_r^2 + \kappa_r^2 + \omega^2/y^2}{\hat{\theta}^2} \right) = N(0, 1) \]
for \(h \downarrow 0\).
Corollary 2  Consider the INR estimator \( \tilde{\text{INR}} = \exp(\tilde{y}) \). For small variance, \( \vartheta^2 \), and conditional on the true \( \log(\text{INR}) \), \( y \), an approximate lower bound for the coefficient of variation of the \( \text{INR} \) estimator is given by

\[
\text{CV} \{ \tilde{\text{INR}} \} \geq y \vartheta \geq \left\{ y^2 \left( \zeta_r^2 + \kappa_r^2 \right) + \omega^2 \right\}^{1/2} \tag{5.22}
\]

Remark: By (5.22), (5.14), and (5.15),

\[
\text{CV} \{ \tilde{\text{INR}} \}
= \left\{ c \log^2(\text{PTR}) \left( \text{Var} \{ \tilde{c} \} / c^2 + \text{Var} \{ c_\lambda \} / c^2 \right) + \omega^2 \right\}^{1/2}
= \left\{ (\text{Var} \{ \tilde{c} \} + \text{Var} \{ c_\lambda \}) \log^2(\text{PTR}) + \omega^2 \right\}^{1/2} \tag{5.23}
\]

where \( \text{PTR} \) is the Prothrombin Time Ratio, \( \log(\text{PTR}) = x - \mu \).

When the interlaboratory variance \( \gamma_i^2 \) is ignored, this expression reduces to

\[
\text{CV} \{ \tilde{\text{INR}} \} \approx \left[ \text{Var} \{ \tilde{\text{ISI}} \} \log^2(\text{PTR}) + \omega^2 \right]^{1/2}
\]

Therefore, the quantity

\[
\text{se}_{\text{eff}}(\tilde{\text{ISI}}) = (\text{Var} \{ \tilde{c} \} + \text{Var} \{ c_\lambda \})^{1/2} \approx c \left( \zeta_r^2 + \kappa_r^2 \right)^{1/2} \tag{5.24}
\]

may be considered an estimate of the effective standard error of the ISI estimate. This quantity can be used for a rough calculation of the minimum CV of INR estimates by the formula

\[
\text{CV} \{ \tilde{\text{INR}} \} \approx \left\{ \text{se}_{\text{eff}}(\tilde{\text{ISI}}) \log^2(\text{PTR}) + \omega^2 \right\}^{1/2} \tag{5.25}
\]

Conventionally, when ISI estimates of new IRPs are reported, only the marginal standard error of the ISI estimate, \( \text{se}(\tilde{\text{ISI}}) = \text{Var}^{1/2} \{ \tilde{c} \} \), is given. It is, however, highly recommended that the effective standard error of the ISI should be reported in order to enable a reasonable estimation of the accuracy of INR estimates derived from that IRP.

Proof:

By theorem 1, \( \log(\tilde{y}) \sim \text{N}(\log(y), \vartheta^2) \). Thus, by theorem 5 in appendix A.2,

\[
\text{E} \{ \tilde{y} \} \approx \exp \left( \log(y) + \vartheta^2 / 2 \right) = y \exp \left( \vartheta^2 / 2 \right)
\]

\[
\text{Var} \{ \tilde{y} \} \approx \exp \left( 2 \log(y) + \vartheta^2 \right) \{ \exp(\vartheta^2) - 1 \}
= y^2 \exp(\vartheta^2) \{ \exp(\vartheta^2) - 1 \}
\]
5.3. DISTRIBUTION OF THE INR ESTIMATOR

Moreover, for $\vartheta$ small, the normal and lognormal distributions closely resemble each other (see e.g. [121]). Thus we may assume that $\tilde{y}$ is approximately normally distributed, and by applying theorem 5 once more, we find that

\[
E\{\text{INR}\} \approx \exp \left[ y \exp \left( \vartheta^2 / 2 \right) + y^2 / 2 \exp \left( \vartheta^2 \right) \left\{ \exp \left( \vartheta^2 \right) - 1 \right\} \right]
\]

\[
\text{Var}\{\text{INR}\} \approx \exp \left[ 2y \exp \left( \vartheta^2 / 2 \right) + y^2 \exp \left( \vartheta^2 \right) \left\{ \exp \left( \vartheta^2 \right) - 1 \right\} \right] \times \left[ \exp \left\{ y^2 \exp \left( \vartheta^2 \right) \left[ \exp \left( \vartheta^2 \right) - 1 \right] \right\} - 1 \right]
\]

Therefore,

\[
\text{CV}^2\{\text{INR}\} = \frac{\text{Var}\{\text{INR}\}}{E^2\{\text{INR}\}}
\]

\[
\approx \exp \left[ y^2 \exp \left( \vartheta^2 \right) \left\{ \exp \left( \vartheta^2 \right) - 1 \right\} \right] - 1
\]

\[
\geq \exp \left( y^2 \vartheta^2 \right) - 1
\]

\[
\geq y^2 \vartheta^2
\]

\[\square\]

5.3.1 A simulation study

The results in section 5.2 for the distribution of the INR estimates are asymptotic, and the following simulation study was performed to evaluate the practical usefulness of the results. The SAS program, which was used to do the simulation, can be found in appendix B.3.

The simulation was based on the model in section 5.2.2 with three thromboplastins $t_1$, $t_2$, and $t_3$. A grid was constructed by assigning the slopes $b_1$, $b_2$, and $b_3$ the values 0.8, 1.0, and 1.2 each, and the inter-laboratory standard deviations $\gamma_1$, $\gamma_2$, and $\gamma_3$ the values 0, 0.05, and 0.1 each. Thus the number of points in the grid was $3^3 \times 3^3 = 729$. For each point, the true log(INR) was sampled from a normal distribution with mean 1 and standard deviation 0.4, which was truncated at log(1.2) and log(8). The number of laboratories participating in each of the three calibrations was assumed to be $k_1 = 7$, $k_2 = 16$, and $k_3 = 11$, respectively. The mean and standard deviation of the true log(PT) of normal persons was assumed to be $\mu = 3$ and $\tau = 0.05$, and the intralaboratory analytical error of log(PT) $\sigma = 0.04$. The number of normal persons used for estimating the mean normal PT was $n = 20$. The values of the parameters were chosen to reflect the values seen in real calibration exercises.
Finally, the intralaboratory standard deviation of the slope estimate was assumed to be log-normally distributed, \( \log(\nu_{il}) \sim N(-4, 0.25^2) \). This assumption was based on examinations of the empirical distributions of estimates of intralaboratory standard deviations from four calibrations [18, 13, 14].

In order to investigate the consistency of the simulated values of \( \log(\hat{y}) \) with the normal distribution, a series of quantile-quantile plots of the empirical distribution of \( \log(\hat{y}) \) against the normal distribution was drawn for a selection of points in the grid. The points were chosen from the edge and the center of the grid, and for each of these points 3000 values of \( \log(\hat{y}) \) were simulated. In all cases, the distribution of \( \log(\hat{y}) \) was consistent with the normal distribution. As an example of a simulation in a point in the center of the grid, see figure 5.2. The simulated values of \( \log(\hat{y}) \) are consistent with a normal distribution, while the values of \( \hat{y} \) are slightly skewed to the right (median 0.9203 vs. mean 0.9236). However, for all practical purposes, such as deriving the CV of INR in corollary 2, the distribution of \( \hat{y} \) may be considered normal.

Next, for each point in the grid, 1000 simulations were performed, and the mean and empirical variance were calculated for both \( \log(\hat{y}) \) and \( \hat{y} \) by taking the average and by the usual formula for the variance, \( \sum (x_i - \bar{x})^2/(n - 1) \). In addition, the asymptotic mean, \( \log(y) \), and the asymptotic variance \( \vartheta^2 \) of \( \log(\hat{y}) \) were calculated.

Thereafter, the resulting 729 data points were investigated. To simplify, we will in the following denote the empirical and asymptotic values of \( \operatorname{Var}[\log(\hat{y}) | y] \) by \( V_e \) and \( V_a \), respectively.

In figure 5.3 \( \log(V_e) \) is plotted against \( \log(V_a) \) and the fitted regression line is drawn. In figure 5.4 the differences between \( \log(V_e) \) and \( \log(V_a) \) are plotted against \( V_a \). The figure indicates that \( \vartheta^2 \) in general slightly underestimates the variance and that the bias is independent of the size of the variance. A probability plot of the differences shows no deviation from a normal distribution with mean 0.0176 and standard deviation 0.0251.

Let \( CV_e \) and \( CV_a \) denote the empirical and asymptotic values of \( \operatorname{CV}(\text{INR}) \), respectively. By using (5.22) we find that the relative difference between \( CV_e \) and \( CV_a \) is

\[
\frac{CV_e - CV_a}{CV_a} \approx \log \left( 1 + \frac{CV_e - CV_a}{CV_a} \right) \\
= \log(CV_e) - \log(CV_a) \\
\approx \log(yV_e) - \log(yV_a) \\
= \log(V_e) - \log(V_a)
\]
5.3. DISTRIBUTION OF THE INR ESTIMATOR

Figure 5.2: Simulation study. Quantile-quantile plot of the empirical distribution of 3000 simulated values of $\log(\tilde{y})$ (left) and $\tilde{y}$ (right) against the normal distribution for a point in the center of the grid.

$\sim N(0.0176, 0.0251^2)$

Thus the 95% prediction interval for the relative difference between $\text{CV}_e$ and $\text{CV}_a$ is approximately $[-3.2\%, 6.7\%]$, and this means that $\text{CV}(\text{INR})$ is well determined by (5.22).

In figure 5.5 the empirical and asymptotic values of $\text{E} \{\log(\tilde{y}) \mid y\}$ ($U_e$ and $U_a$, respectively) are plotted against each others, and in figure 5.6 the differences between $U_e$ and $U_a$ are plotted against the true INR. This plot indicates that the agreement between the empirical and asymptotic means is good, but that $\log(\tilde{y})$ slightly underestimates the mean of $\log(\tilde{y})$. Moreover, for INR $\in [2, 6]$, the difference between $U_e$ and $U_a$ seems not to depend on the size of INR. A probability plot shows that for INR in this interval the differences can be roughly approximated by a normal distribution with mean $5.55 \times 10^{-3}$ and standard deviation $6.02 \times 10^{-3}$. 
Figure 5.3: Simulation study. Scatterplot of $\log(V_e)$ against $\log(V_a)$ with fitted regression line (solid), $\log(V_e) = 0.0434 + 1.0122 \log(V_a)$, and identity line (dashed).

Figure 5.4: Simulation study. Scatterplot of differences between $\log(V_e)$ and $\log(V_a)$ against $V_a$. 
Figure 5.5: Simulation study. Scatterplot of $U_e$ against $U_a$ with fitted regression line (solid), $U_e = 0.0049 + 1.0046U_a$ and identity line (dashed).

Figure 5.6: Simulation study. Scatterplot of differences between $U_e$ and $U_a$ against true INR. Note that the horizontal axis is log scaled.
All in all, the simulation study is consistent with the statements in theorem 1. The result in theorem 1 should, however, not be used to estimate the bias of INR estimates. To see this, note that according to theorem 1 and theorem 5 in appendix A.2, the asymptotic mean of \( \tilde{y} \) is

\[
E \{ \tilde{y} \} \approx y \exp \left( \frac{\vartheta^2}{2} \right) \approx y \left( 1 + \frac{\vartheta^2}{2} \right)
\]

Therefore, the asymptotic relative bias of \( \tilde{y} \) is

\[
E \left\{ \frac{\tilde{y} - y}{y} \right\} = E \left\{ \frac{\tilde{y}}{y} - 1 \right\} \approx y \left( 1 + \frac{\vartheta^2}{2} \right) / y - 1 = \frac{\vartheta^2}{2}
\]

which (in the first generation of IRPs) is of the magnitude

\[
\frac{1}{2} \vartheta^2 = \frac{1}{2} \left\{ \zeta_i^2 + \kappa_i^2 + c^2 / y^2 \left( \sigma^2 + \tau^2 / n \right) \right\}
\]

\[
= \frac{1}{2} \left\{ 0.02^2 + 0.05^2 + 1^2 / 1^2 \left( 0.04^2 + 0.05^2 / 20 \right) \right\} \approx 2.3 \times 10^{-3}.
\]

Moreover,

\[
\frac{\tilde{y} - y}{y} \approx \log \left( 1 + \frac{\tilde{y} - y}{y} \right) = \log (\tilde{y}) - \log (y) = U_e - U_a
\]

Thus the model uncertainty is of the same magnitude as the asymptotic relative bias.

### 5.4 The interlaboratory variation

The results in section 5.3 show that the precision of the INR estimate is highly dependent on the interlaboratory variance, \( \gamma_i^2 \). Thus it is important to obtain a reliable estimate of this parameter. In the following theorem, we will derive an unbiased method-of-moments estimator.

**Theorem 3** Assume the model of section 5.2 and consider a fixed thromboplastin \( i \in \{1, \ldots, r \} \). Suppose that \( \tilde{\nu}_l^2 \) is an unbiased estimator of \( \nu_{il}^2 \), \( l = 1, \ldots, k_i \). Then an unbiased method-of-moments estimator of the interlaboratory variance \( \gamma_i^2 \) is

\[
\tilde{\gamma}_i^2 = s_i^2 - \tilde{\nu}_i^2
\]

where

\[
s_i^2 = \frac{1}{k_i - 1} \sum_{l=1}^{k_i} (\tilde{b}_{il} - \tilde{b}_i)^2
\]

\[
\tilde{\nu}_i^2 = \frac{1}{k_i} \sum_{l=1}^{k_i} \tilde{\nu}_{il}^2
\]
Remark: The term $s_i$ is commonly called the “interlaboratory standard deviation of the slope estimate” [13, 17, 18]. It is, however, important to note that this term includes both the between- and the within-laboratory variation and therefore should be called the total standard deviation.

Suppose that the correlation between calibrations within the same laboratory is zero, i.e. $\rho_{i\neq i'} = 0$ for $i \neq i', i, i' = 1, \ldots, r$. Then the expressions for $\zeta^2_r$ and $\kappa^2_r$ reduce to

$$\zeta^2_r = \sum_{i=1}^{r} \frac{\gamma_i^2}{b_i^2 k_i} + \nu_i^2,$$

$$\kappa^2_r = \sum_{i=1}^{r} \frac{\gamma_i^2}{b_i^2},$$

and in this case a natural estimator of the effective standard error of ISI is

$$\tilde{s}_{\text{eff}}(\tilde{\text{ISI}}) = \tilde{c} \left\{ \sum_{i=1}^{r} \frac{s_i^2}{b_i^2 k_i} + \sum_{i=1}^{r} \frac{s_i^2 - \tilde{\nu}_i^2}{b_i^2} \right\}^{1/2}.$$  \hspace{1cm} (5.26)

This formula clearly demonstrates that the effective standard error has two components; one, which can be reduced by increasing the number of participating laboratories, $k_i$, and another, which is due to interlaboratory variation, and which cannot be reduced by increasing the number of laboratories. Traditionally, however, only the marginal standard error of the ISI,

$$\tilde{s}_e (\tilde{\text{ISI}}) = \tilde{c} \left\{ \sum_{i=1}^{r} \frac{s_i^2}{b_i^2 k_i} \right\}^{1/2},$$

is reported, and this may lead to serious overestimation of the precision of INR estimates.

Proof: The expectation of the sum of squares is

$$E \left\{ \sum_{i=1}^{k_i} (\tilde{b}_{il} - \tilde{b}_i)^2 \right\}$$

$$= \sum_{l=1}^{k_i} E \left\{ (\tilde{b}_{il} - \tilde{b}_l)^2 \right\}$$

$$= \sum_{l=1}^{k_i} \left\{ \gamma_l (U_{il} - \tilde{U}_l) + \left( \nu_{il} V_{il} - k_i^{-1} \sum_{i'=1}^{k_i} \nu_{i'l'} V_{i'l'} \right)^2 \right\}$$

$$= \sum_{l=1}^{k_i} \text{Var} \left\{ \gamma_l (U_{il} - \tilde{U}_l) \right\} + \sum_{l=1}^{k_i} \text{Var} \left\{ \nu_{il} V_{il} - k_i^{-1} \sum_{i'=1}^{k_i} \nu_{i'l'} V_{i'l'} \right\}.$$
\[ \gamma_i^2 \sum_{l=1}^{k_i} \left[ \text{Var} \left\{ \frac{k_i - 1}{k_i} U_{il} \right\} + \text{Var} \left\{ \frac{k_i - 1}{k_i} \sum_{l' \neq l} V_{il'} \right\} \right] + \sum_{l=1}^{k_i} \nu_{il}^2 \left[ \text{Var} \left\{ \frac{k_i - 1}{k_i} V_{il} \right\} + \text{Var} \left\{ \frac{k_i - 1}{k_i} \sum_{l' \neq l} V_{il'} \right\} \right] \]

\[ = (k_i - 1) \gamma_i^2 + (k_i - 1) \sum_{l=1}^{k_i} \nu_{il}^2 \]

Now, equating the sum of squares with its expectation yields the desired result. \(\Box\)

According to Fuller [117] p. 34 the approximate distribution of the variance of the slope estimate within the \(l\)’th laboratory is given by

\[ \tilde{\nu}_{il}^2 \sim \nu_{il}^2 \chi^2(f_{il}) / f_{il} \]

where \(f_{il}\) is the number of points in the calibration minus 2. This allows us to evaluate the statistical significance of the inter-laboratory variance according to the following theorem.

**Theorem 4** For a fixed thromboplastin \(i \in \{1, \ldots, r\}\), assume that the estimates of the within-laboratory variances of the slope estimates, \(\tilde{\nu}_{il}^2\) are scaled \(\chi^2\) distributed,

\[ \tilde{\nu}_{il}^2 \sim \nu_{il}^2 \chi^2(f_{il}) / f_{il}, \]

mutually independent, and independent of the slope estimates \(\tilde{b}_{il}\).

1. The full model: The maximum likelihood estimates of the inter-laboratory variance, \(\gamma_i^2\), the common slope, \(b_i\), and the within-laboratory variance of the slope estimates, \(\nu_{il}^2\), are given by the following algorithm:

   (a) Initial estimates: Let

   \[ b_i := \tilde{b}_i = \frac{1}{k_i} \sum_{l=1}^{k_i} \tilde{b}_{il} \]

   \[ \gamma_i^2 := s_i^2 + \tilde{\nu}_{i}^2 \]

   \[ = \frac{1}{k_i - 1} \sum_{l=1}^{k_i} (\tilde{b}_{il} - \tilde{b}_i)^2 - \frac{1}{k_i} \sum_{l=1}^{k_i} \nu_{il}^2 \]
5.4. THE INTERLABORATORY VARIATION

(b) Let

\[ \alpha_{il} := \frac{\gamma_i^2}{(\tilde{b}_{il} - b_i)^2}, \quad l = 1, \ldots, k_i \]

\[ \beta_{il} := \frac{\nu_{il}^2}{(\tilde{b}_{il} - b_i)^2}, \quad l = 1, \ldots, k_i \]

(c) Solve the cubic equation

\[ a_{il}^3 - (\alpha_{il} + 2) a_{il}^2 + [(f_{il} + 2) + f_{il} \beta_{il} + 1] a_{il} - (f_{il} + 1) \alpha_{il} = 0 \]

for \( a_{il}, \quad l = 1, \ldots, k_i \).

(d) Calculate

\[ \gamma_i^2 := \frac{\sum_{l=1}^{k_i} a_{il} \left[ (\tilde{b}_{il} - b_i)^2 + f_{il} \nu_{il}^2 \right]}{\sum_{l=1}^{k_i} f_{il} + k_i}, \quad b_i := \frac{\sum_{l=1}^{k_i} a_{il} \tilde{b}_{il}}{\sum_{l=1}^{k_i} a_{il}} \]

Repeat steps 1b-1d until convergence. The ML estimates of the intra-laboratory variance of the slope estimates are given by

\[ \nu_{il}^2 = \frac{\gamma_i^2 (1 - a_{il})}{a_{il}}, \quad l = 1, \ldots, k_i \]

2. The reduced model: Under the hypothesis \( \gamma_i = 0 \) the ML estimates of the common slope, \( b_i \), and the variance of the intra-laboratory slope estimates, \( \nu_{il}^2 \), are given by the following algorithm:

(a) Initial estimate: Let

\[ \nu_{il}^2 := \tilde{\nu}_{il}^2, \quad l = 1, \ldots, k_i \]

(b) Calculate

\[ b_i := \frac{\sum_{l=1}^{k_i} \nu_{il}^{-2} \tilde{b}_{il}}{\sum_{l=1}^{k_i} \nu_{il}^{-2}} \]

\[ \nu_{il}^2 := \frac{(\tilde{b}_{il} - b_i)^2 + f_{il} \tilde{\nu}_{il}^2}{f_{il} + 1}, \quad l = 1, \ldots, k_i \]
Repeat step 2b until convergence.

Remark:

- Let $\hat{b}_i$, $\hat{\gamma}_i^2$, and $\hat{\nu}_i^2$, $l = 1, \ldots, k_i$ be ML estimates in the full model. Vangel & Rukhin [122] showed (in theorem 3 of that paper) that $\hat{\nu}_i^2 \geq \tilde{\nu}_i^2$ if and only if $|\tilde{b}_i - \hat{b}_i| \geq (\hat{\gamma}_i^2 + \hat{\nu}_i^2)^{1/2}$, i.e. if and only if the $l$th laboratory in some sense is an outlier. They also demonstrated (theorem 2) that for each $l$, the cubic equation in the iterative procedure for the full model has either one or three roots in the unit interval and that a necessary condition for the existence of three roots in $[0, 1]$ is that

$$
(\tilde{b}_i - b_i)^2 \geq \max \left( \frac{\hat{\gamma}_i^2}{q_{il}}, \frac{\hat{\nu}_i^2}{h_{il}} \right)
$$

where

$$
q_{il} = \frac{8}{27 (f_{il} + 1)} + O \left( f_{il}^{-2} \right)
$$

and

$$
h_{il} = \frac{1}{27 (f_{il} + 1)} + O \left( f_{il}^{-2} \right).
$$

- A test for the hypothesis $H$: $\gamma_i = 0$ can be constructed by utilizing that under $H$ the deviance (i.e. twice the log likelihood ratio) between the full and the reduced model is asymptotically $\chi^2$ distributed with one degree of freedom.

Proof:

1: If each laboratory had performed a number of independent estimations of the slope, we would have the setup of Vangel & Rukhin [122]. In that case, since the average and the sum of squares within each laboratory are sufficient statistics for the unknown parameters, we might as well only have observed these independent statistics. This is similar to our case where we have observed the slope estimate within each laboratory and an approximately unbiased and $\chi^2$ distributed estimate of the variance of the average. Thus we can use the result by Vangel & Rukhin, theorem 1, which guarantees that the stated iteration procedure converges to a stationary value of the likelihood, and that the iteration is monotone in the sense that the likelihood is increased at each step. This completes the proof for the full model.
2: Minus twice the log likelihood is
\[
l = \sum_{i=1}^{k_i} (f_{il} + 1) \log(\nu_{il}^2) + \left[ (\tilde{b}_{il} - b_i) + f_{il} \tilde{\nu}_{il}^2 \right]
\]
The derivatives are
\[
\frac{\partial l}{\partial b_i} = -2 \sum_{i=1}^{k_i} \nu_{il}^2 \left( \tilde{b}_{il} - b_i \right)
\]
\[
\frac{\partial l}{\partial \nu_{il}^2} = \frac{f_{il} + 1}{\nu_{il}^2} - \left( \frac{(\tilde{b}_{il} - b_i)^2 + f_{il} \tilde{\nu}_{il}^2}{\nu_{il}^4} \right), \quad l = 1, \ldots, k_i
\]
Now, solving the likelihood equations
\[
\frac{\partial l}{\partial b_i} = 0, \quad \frac{\partial l}{\partial \nu_{il}^2} = 0, \quad l = 1, \ldots, k_i
\]
yields the desired result. \[\square\]

5.5 Example: Estimates derived from RBT/90

The international reference preparation RBT/90 is a second generation IRP with ancestors RBT/79 and IRP 67/40. IRP 67/40 is considered the 0’th generation (see figure 5.7). In the following we will calculate the CV of INR-estimates based on these IRPs.

5.5.1 \( t_1 \) versus \( t_0 \): Calibration of RBT/79 against IRP 67/40

The calibration of RBT/79 against the primary IRP 67/40 is described by Hermans et al. [13]. Seven laboratories \( k_1 = 7 \) participated in this calibration, and each laboratory performed PT determinations over 6–10 days on 12–20 normal plasmas and 36–60 abnormal plasmas. Table 5.1 shows the estimates from the article. In accordance with our notation, \( l \) is the laboratory number, \( f_{il} \) is the number of points minus two, \( \tilde{b}_{il} \) is the slope estimate from the \( l \)’th laboratory, and \( \tilde{\nu}_{il} \) is the estimated standard deviation of \( \tilde{b}_{il} \).
$t_0$: IRP 67/40

$\tilde{b}_1 = 1.413$

$\tilde{b}_2 = 0.732$

$t_1$: RBT/79

$t_2$: RBT/90

Figure 5.7: Overview of the ancestors of the IRP RBT/90 with relating slope estimates

Table 5.1: Results from the calibration of RBT/79 against IRP 67/40.
5.5. EXAMPLE: ESTIMATES DERIVED FROM RBT/90

The average of the slope estimates is $\hat{b}_1 = 1.413$, and the total variance of the slope estimates is

$$s_1^2 = \frac{1}{(k_1 - 1)} \sum_{l=1}^{k_1} (\hat{b}_{1l} - \hat{b}_1)^2 = 0.0963^2$$

The average estimated variance of the slope estimates is

$$\hat{\nu}_1^2 = \frac{1}{k_1} \sum_{l=1}^{k_1} \hat{\nu}_{1l}^2 = 0.0384^2$$

Now, according to theorem 3, an unbiased method-of-moments estimator of the interlaboratory variance, $\gamma_1^2$, is

$$\hat{\gamma}_1^2 = s_1^2 - \hat{\nu}_1^2 = 0.0884^2$$

Next, it is relevant to ask if the interlaboratory variance is significant. By using theorem 4 and the SAS program in appendix B.2, we find that minus twice the maximum log likelihood for the full model is -1681, and -1041 for the reduced model. Thus the deviance is 640 in a $\chi^2$ distribution with one degree of freedom, i.e. the interlaboratory variance is highly significant! The ML estimator of $\gamma_1$ is 0.0891, which is close to the unbiased estimator. The ML estimator of the common mean slope $b_1$ is 1.413, which is the same as the average of the slope estimates $\hat{b}_1$. There is practically no difference between the individual estimates $\hat{\nu}_{1l}$ and the ML estimates $\hat{\nu}_{1l}$ obtained by running the program. This was to be expected, since none of the laboratories are outliers.

The procedure converged nicely, albeit a bit slowly, after approximately 300 and 25 iterations for the full and the reduced model, respectively.

Now we can calculate estimates of $\zeta_1$ and $\kappa_1$ by the formulae (5.10) and (5.11),

$$\hat{\zeta}_1^2 = \frac{\hat{\gamma}_1^2 + \hat{\nu}_1^2}{b_1^2 k_1} = \frac{s_1^2}{b_1^2 k_1}$$

$$= \frac{0.0963^2}{1.413^2 \cdot 7} = 0.0258^2$$

$$\hat{\kappa}_1^2 = \frac{\hat{\gamma}_1^2}{b_1^2} = \frac{0.0884^2}{1.413^2} = 0.0625^2$$

In this calibration exercise, since the reference thromboplastin preparation is the IRP 67/40, the slope estimates are equal to the ISI estimates. Thus the CVs of the common ISI estimate, $\hat{c}_1$, and the true laboratory-specific ISI,
ACCURACY OF INR ESTIMATES

c_\lambda, for RBT/79 are approximately 2.58% and 6.25%, respectively. By (5.24) the effective standard error of the ISI estimate is

$$c_1 \sqrt{\xi_1^2 + \kappa_1^2} \approx 1.413 \sqrt{0.0258^2 + 0.0625^2} = 0.0956$$

Thus, due to the large interlaboratory variation, the effective standard error of the ISI estimate is very close to the total standard deviation, $s_1 = 0.0963$. In comparison the traditionally reported marginal standard error of the ISI (see section 5.4) is

$$\tilde{s}_e (\tilde{ISI}) = \tilde{c} \left( \frac{s_1^2}{b_1^2 k_1} \right)^{1/2} = \frac{s_1}{k_1^{1/2}} = 0.0364$$

In order to find a minimum estimate of the variance of INR, assume that for a given patient, his INR is determined directly by using RBT/79 according to the WHO recommendations. Moreover, we assume that the mean normal PT is determined exactly such that $\tilde{\mu} = \mu$. If the long-term within-laboratory analytical CV is $\sigma = 0.03$, and the true log(INR) is $y = \log(2.5)$, then, according to corollary 2, the estimate of the CV of the INR estimate is

$$\tilde{CV} \{\tilde{\text{INR}}\} \approx 7.5\%$$

For the true INR being 2 and 4, the corresponding estimates are 6.3% and 10.3%, respectively. Calculations by means of the effective standard error of the ISI using (5.25) give exactly the same results.

Note that the estimates given are minimal. In routine use, the CV will be larger, since the working thromboplastin will have been calibrated against at least one other thromboplastin, and the mean normal PT will have been determined with some imprecision. Moreover, the long-term analytical within-laboratory CV is probably larger than the 3% that we have assumed here. In a previous study [116] we found indications that it was closer to 8%.

The variance of the mean normal log PT estimate depends on the variance of the log PTs in the population of normal plasmas. By using the results in table 3 in [13], the total CV of the PTs of the normal plasmas is found to be approximately 6.9%. This is also approximately the square root of the sum of the variance of the true log PTs of normal plasmas $\tau^2$ and the within-laboratory analytical CV, $\sigma^2$. Assuming that the mean normal PT is determined by using 20 normal plasmas ($n = 20$), and that $\sigma = 0.05$, then (see (5.12)) $\omega \approx 1.413 \{0.05^2 (1 - 1/20) + 0.063^2/20\}^{1/2} = 0.072$.

Under these more realistic assumptions, the CV of INR estimates measured directly by RBT/79 according to WHO recommendations is 8.6% and 11.8% for the true INR being 2 and 4, respectively. This implies that the 95% prediction interval for the INR estimates is [1.69, 2.37] and [3.17, 5.04], respectively.
5.5.2  $t_2$ versus $t_1$: Calibration of RBT/90 against RBT/79

RBT/90 was calibrated against RBT/79 by $k_2 = 20$ laboratories [17]. We have no information on laboratory-specific slope estimates, so in order to obtain minimum estimates of the variance, it is assumed that the interlaboratory variance is 0. Moreover, it is assumed that the correlation between calibrations within the same laboratory is 0. This is, however, probably not the case and this should be investigated further. In case of positive correlations, the variance estimates obtained in this way are minima.

The relevant estimates for RBT/90 against RBT/79 are: $\tilde{b}_2 = 0.732$, and $s_2 = 0.021$. Accordingly, the ISI estimate of RBT/90 is $\tilde{c}_2 = 1.035$. The estimates of $\kappa_2$ and $\zeta_2$ are $\tilde{\kappa}_2 \geq \tilde{\kappa}_1 = 0.0625$, and $\tilde{\zeta}_2 \geq \tilde{\zeta}_1 = 0.0258$.

By insertion the effective standard error of the ISI estimate is found to be greater than $1.035 (0.0625^2 + 0.0258^2)^{1/2} = 0.070$. In comparison the standard error of the ISI estimate was reported as 0.027 [17]. If this estimate is used in calculations of the uncertainty of INR estimates, it will result in gross overestimation of the precision of INR.
6

Discussion

6.1 Precision of home coagulometers

The purpose of this substudy was to investigate over time the precision of INR values obtained by patient self-testing using a home coagulometer in a self-managed OAT regimen.

Macik [123] discussed some of the problems in validating capillary whole blood coagulometers. “New technology is typically compared to a ‘gold standard’ and is deemed acceptable or unacceptable based on this performance. Since a true standard does not exist for . . . coagulation assays, professionals need to establish the criteria by which to judge the validity and merit of a point-of-care coagulation system. Key items that will influence the development of these criteria include (1) realistic performance expectations of the PT and APTT assays; (2) attention to preanalytical handling of the blood sample; (3) understanding the difference between whole blood and plasma tests; (4) patient populations defined for point-of-care testing; (5) personnel requirements; and (6) special point-of-care considerations. . . .

Although the methods are similar and the names the same, a whole blood PT or APTT is not identical to its plasma counterpart. Red blood cells, white blood cells, and platelets all contribute to the process of coagulation. In fact, the ability to test whole blood may be a truer representation of the in vivo coagulation process. Since reliable whole blood assays are relatively new, the unique ability of these to assays to monitor the coagulation process has not been adequately studied. Unfortunately, most manufacturers of whole blood instruments have adjusted the whole blood clotting times to mimic those achieved with plasma, instead of exploring the unique properties of whole blood monitoring.”

It is possible, and even plausible, that the whole blood test is more phys-
6. DISCUSSION

...logical than the plasma method, and that this should be explored, as suggested by Macik. At present there is, however, no reason to believe that the INR system is inadequate for whole blood testing, and thus whole blood instruments should be evaluated with respect to INR.

**Precision**

In our study we found an almost complete agreement between INR results obtained from patients’ own files and downloaded results from home coagulometer memories.

Other studies [4, 112] reporting on CoaguChek coagulometers found within-run analytical CVs of 3.1% and 4% using venous blood as test substance, and the lower precision in our study (5.4%) probably reflects the additional source of variation from different finger punctures. In an earlier study [116] we found a long-term total analytical CV of approximately 10%, and thus the long-term between-day analytical CV is approximately 8.4%.

The precision of the CoaguChek coagulometers should be held against that of clinical chemistry laboratories and the biological variation. Based on data from the Danish Institute for External Quality Assurance in Hospital Laboratories (personal communication, data not shown) the inter-laboratory CV of approximately 65 quality controlled Danish hospital laboratories was 8.3% (CI 7.9% to 8.8%) in 1996-1997.

Based on measurements on pooled control plasma (approximately 2.5 INR) in the local quality controlled hospital laboratory (data not shown), we found a within-day analytical CV of 3.6% and a long-term between-day analytical CV of 6.3%, giving a long-term total analytical CV of 7.3%. In our earlier study [116] we found that the long-term total analytical CV of the local hospital laboratory was approximately 8%, and thus the long-term between-day analytical CV was approximately 7.1%. The reason for the slightly larger estimates in our earlier study may be that the analytical precision of laboratory analyses is smaller for individual patient plasma specimens than for pooled frozen plasma.

In our earlier study [116] we also estimated that the long-term within-subject biological CV in self-managing patients was approximately 19% (and approximately 30% in a retrospectively selected control group of patients on conventional treatment).

In this perspective the precision of the home coagulometer is acceptable.
6.1. PRECISION OF HOME COAGULOMETERS

Trueness

Since the CV of INR estimated from IRPs is large (see the example in section 5.5), no gold standard exists for INR. In particular, routine laboratory INR estimates cannot be used as reference values (see page 30 and Kaatz et al. [4]). Thus it is unclear how the trueness of coagulometers should be evaluated. The substudy was not planned to address the question of the trueness of the home coagulometers, and the material is not strong enough to draw any firm conclusions regarding the trueness of neither home coagulometers nor laboratories.

Simple calculations based on estimates of laboratory and CoaguChek variation show that the deviations between laboratories and CoaguChek coagulometers observed in our study are not unlikely (see page 55). In August 1997, where we compared results from the home coagulometers with those of two different laboratories, the deviation between the home coagulometers and lab B was not larger than between the two laboratories, and no deviations were larger than we would expect from our estimates of laboratory and home coagulometer variation. Other studies [4, 112] on CoaguChek coagulometers have found acceptable agreement with the reference method recommended by the WHO, and inaccuracies of a magnitude comparable to those of routine laboratory methods. Altogether, in the present study there is no support for claiming that the home coagulometers are inaccurate.

As noted on page 39 McCurdy & White [105] found that CoaguChek underestimation increased with the level of INR. Figure 2 in their paper is a plot of differences between monitor and laboratory versus laboratory INR. Apparently, there is a trend towards lower mean difference with increasing INR values. The plot is, however, misleading, since the laboratory INR values are measured with error. In appendix A.1 it is shown how such a plot will display a negative trend, even though both the monitor and laboratory values are unbiased estimates of the true INR. In this case, the slope $a$ of the line relating the monitor INRs to the laboratory INRs is approximately one plus the slope $b$ of the line relating the differences to the laboratory INRs. The intercepts of the two lines may be assumed to be equal, $c$. McCurdy & White found that $a = 0.68$ and $c = 0.91$. Thus, $b = 0.68 - 1 = -0.32$. This agrees very well with their observation that “the monitor measured on average 0.3 INR units high at a criterion [i.e. laboratory] value of 2.0 units and less than 0.1 units low at a criterion standard value of 3.0 INR units. At a criterion standard value of 4.5 INR units, the monitor measured on average 0.5 units low.” Of course this observation is true, but it is irrelevant. What is relevant is whether the difference depends on the true value, which, unfortunately, is unknown. In any case the authors do wrong in concluding
that “the values obtained with the monitor tended to differ systematically from those obtained from the reference laboratory”. This can not be seen from the data in their paper.

The question has arisen whether portable coagulometers like CoaguChek should be equipped with the possibility of being locally calibrated. It seems, however, that this is caused by a misunderstanding of the WHO guidelines, which recommend \textit{local-system calibration} in the sense that every local PT system should be calibrated. The CoaguChek System, with its test carriers and code chips, is indeed a local PT system. In this sense the only difference between CoaguChek and laboratory PT systems is that the former is portable, whereas the latter is usually not. The concept of portability comprises that the act of moving the coagulometer to another place, perhaps with different climate etc., has no influence on its INR estimates. Therefore, a portable coagulometer (or a sufficient part of it, i.e. the code chip) may be validly calibrated at the manufacturer’s premises, and no (local or other) calibration should take place afterwards.

6.2 Validity of the INR standard

The primary purpose of this substudy was to examine the relations between the Nycotest, Hepato Quick and CRM 149S methods, with special emphasis on the hypotheses $H_1$ and $H_2$ as specified on page 3. Secondarily we wished to see whether there was a clinically significant difference in INR between Hepato Quick and Nycotest.

We introduced a new method, which in a single plot gives an overall impression of the consistency of calibration data with the hypotheses $H_1$ and $H_2$. In this plot the orthogonal regression line estimated from coumadin log PTs is plotted along with its 95\% confidence interval and 95\% prediction intervals for the log PTs. Moreover, the mean of the normal log PTs with its 95\% confidence ellipse is depicted. This plot visualizes deviations from the assumption of linearity of the log PTs of coumadin plasmas, and outliers can easily be spotted. Moreover, the statistical significance of deviations from $H_2$ can be spotted by comparison of the confidence ellipse with the size of the confidence interval of the line, and the clinical significance can be judged from the distance between the point corresponding to the MNPT and the regression line.

We found that the hypotheses $H_1$ and $H_2$ were consistent with the data for all comparisons between methods. If the manufacturer can obtain a relation between CoaguChek and Hepato Quick, such that $H_1$ and $H_2$ are valid, it can, therefore, be expected that $H_1$ and $H_2$ are also valid for the relation
6.2. VALIDITY OF THE INR STANDARD

between CoaguChek and Nycotest. Moreover, since the data for comparison of Hepato Quick with CRM 149S were consistent with $H_1$ and $H_2$, it is fair to claim that accuracy of CoaguChek can be achieved, at least when CRM 149S is the reference standard.

We found that the difference in INR estimates was 4–5% between Nycotest and CRM 149S, and 6–7% between Hepato Quick and CRM 149S. Thus the difference between Hepato Quick and Nycotest was approximately 1–3%. The differences were not statistically significant. In view of the typical therapeutic target range of 2–3 INR, these differences are considered to be of no serious clinical importance neither.

A priori it was expected that the three different measurements by CRM 149S, namely CRM$_1$, CRM$_2$ and CRM$_3$, “easily” would satisfy the linearity hypotheses $H_1$ and $H_2$. Moreover, we expected that the estimated slopes of the lines relating the three variables would be very close to 1. We found, however, that $H_1$ was inconsistent with the data from the comparison of CRM$_1$ with CRM$_2$, and CRM$_1$ with CRM$_3$. Moreover, the slopes of the lines relating CRM$_1$ to CRM$_2$, and CRM$_1$ to CRM$_3$ were statistically significantly smaller than 1, if two outlying observations were excluded from the analysis (see table 4.2 on page 62). $H_1$ was only consistent with the data from the comparison of CRM$_2$ with CRM$_3$, and only the slope of the line relating CRM$_2$ to CRM$_3$ was close to 1. The reason for these observations may be that the measurements of CRM$_2$ and CRM$_3$ were performed on the same day and two months after the measurement of CRM$_1$. Frozen aliquots of the same plasma were used for all measurements, and it is possible that factor V has degraded in the time between the two sessions. Since the CRM 149S is plain, and thus contains no factor V, this will cause a prolongation of the PT with time. Whether this will cause a change in the ISI depends on the relative prolongations of abnormal and normal PTs. In our study the abnormal PTs were relatively more prolonged than the normals, and the ISI, therefore, decreased. Poller et al. [90], in a study using a plain human (ISI = 0.95) and a plain rabbit (ISI = 1.67) thromboplastin, found that with the human thromboplastin the PT was prolonged by freezing at -40 C for 7 days for both normal and coumarin plasma. With rabbit thromboplastin no significant changes were found. Whether freezing has any effect on the ISI is, however, unclear, but our results indicate that if plasma frozen for a period of two months is used to estimate the ISI of a plain reagent, this may cause the ISI to be underestimated. We are unable to investigate the effect of freezing on PT determinations by Nycotest, since Skejby Sygehus changed reagent between the measurements on fresh and frozen plasmas.

In this substudy and in other similar calibration studies, where only single measurements are performed by each method, it is not possible to provide
a p-value for the linear hypothesis \textit{per se}, since any formal test will hinge on the comparison of the linear model with another more extensive model. Thus, it would be necessary subjectively to choose a more extensive model for comparison, e.g. a polynomial model, and in any case the resulting p-value would have a subjective element. Moreover, in the case of a false hypothesis, the p-value is highly dependent on the number of measurements performed.

It is our view that the examination of the linearity assumption is better performed informally by an experienced statistician, who can do various model checks by various tools such as residual analysis. The model behind the ISI determinations is the classical errors-in-variables model [118, 119]. When the variances of the measurement errors on both axes are equal (as is commonly assumed in ISI calibrations), the procedure for estimating the parameters in this model is called \textit{orthogonal regression}. In this connection it is noteworthy that the BLUEs of the residual and of the true value of the x-variable correspond to the residual and to the explanatory variable, respectively, in an ordinary linear regression, see also Fuller [117]. Draper & Smith [124] give a comprehensive review of techniques for model checking in the linear regression model.

Suppose that the reason for the scatter about the orthogonal regression line is solely analytical error. Then one way to enable a more formal and objective test of \( H_1 \) would be to perform double measurements of each plasma sample by each method. It is, however, important that such double measurements reflect the long-term variation within each method and this will require that the same sample will have the possibility of being measured on two different days according to a carefully planned design. If double measurements are not possible, then at least the measurements should be performed over many days in order to reflect the long-term analytical variation of the methods.

Another major advantage of double measurements is the possibility of estimating the variance of the measurement error for each method. This would enable a better estimation of the line, since it would not be necessary to make the dubious assumption of known ratio between variances of measurement errors. This assumption is dubious, since it may be expected that the methods have different long-term analytical errors. Like in other similar studies, we have, however, no means for investigating this assumption.

The presence of outliers constitutes a delicate problem. The practice of excluding patients because they are outside some prespecified interval may prove dangerous, since the resulting ISI estimate (or at least its confidence limits) will not reflect the true variation in the data material. Thus, the precision of the ISI estimate will be overestimated. Moreover, since the value on the y-axis is considered the reference value, this practice will tend to exclude
points lying above the regression line for high y-values and below the line for small y-values. The ISI will, therefore, be systematically underestimated. In the calibration exercise of Hermans et al. (see page 25) almost 15% of the observations were excluded, and the practice can therefore be expected to have a large impact.

In our study we observed two outliers in the comparison between CRM$_1$ on the one side and CRM$_2$ and CRM$_3$ on the other. In many cases the consequence of an exclusion will be a lower se of the slope estimate, and perhaps a change in the slope estimate as well. In such a case the decision about exclusion should ideally be based on objective evidence, such as instrument failure, since otherwise the estimate of the se will not reflect the reality but lead to erroneously high confidence in the resulting INR estimates. Also in this case, true double measurements would help solving the problem.

In the present study we also observed that 4-6 log PTs of abnormal plasmas were close to the normal range. This is formally a violation of the assumption of normality of abnormal true log PTs, which underlies the derivation of the estimates of the variance of the slope and intercept estimates. As long as linearity is fulfilled, this violation has, however, no influence on the slope and intercept estimates themselves, although the outlying observations may be influential.

The possibility of significant inter-laboratory variation, where each laboratory in fact has its own true slope, presents another difficulty. Estimation of the se of the average slope without adjusting for this may result in serious underestimation of the se of the ISI estimate (see chapter 5).

Having accepted that the data are consistent with both the hypotheses $H_1$ and $H_2$, the logical next step is to incorporate the information from both the normal and abnormal plasmas into a common slope estimate, which should be considered the final result of the calibration exercise. Since the normals and abnormals have different distributions, it is not immediately clear that it is appropriate to estimate the slope by using all the data on the same footing. We are, however, not aware of any method, which takes this into consideration, and more work is needed in order to develop a rational procedure for estimation in this case.

Suppose that $H_1$ is rejected for a pair of thromboplastins, a reference method and a new candidate method. Then the candidate method should be rejected, provided that the consistency of the reference method with $H_1$ has been demonstrated with a high degree of certainty in previous calibrations against other reference methods. If, however, the data are consistent with $H_1$, but $H_2$ is rejected, there are the possibilities of either using Tomenson’s correction, or rejecting the candidate method. The use of Tomenson’s correction will, however, introduce a new parameter, e.g. the distance be-
tween the mean normal log PT and the line relating the PTs of abnormal plasmas, into the INR system.

The idea of normalizing with the MNPT stems from a time when it was thought that a sensible standard for PT could be obtained by using dilution curves (see page 7). From a statistical point of view, however, the MNPT introduces a lot of unnecessary uncertainty, and it can be seriously doubted, whether the assumption of the MNPT lying on the line through patients’ plasma is ever satisfied. Of course, it is not possible to prove that this assumption is true or untrue, but it seems a priori unlikely. A statistical test is highly dependent on the size of the dataset used, and the p-value obtained has no implications for the clinical significance of deviations from the assumption. Therefore, Tomenson’s correction should always be performed, no matter what result the test gives.

The use of a PT ratio may stabilize some of the effects of preanalytical variation (as seen when plasma samples have been frozen), but it may at the same time reduce the sensitivity of the test. It should therefore be investigated whether satisfactory results could be obtained by abolishing the MNPT and giving the relationship between two thromboplastin preparations by the slope and intercept. As long as the INR is perceived as the PT ratio that would have been obtained, had the IRP 67/40 been used, it is still possible to retain the INR standard with this approach.

6.3 Accuracy of INR estimates

Tomenson [45] emphasized that “once it has been demonstrated that the calibration model is a good description of the data, the most important parameter is not the slope of the calibration line but the variation about the line. This is the factor which will determine whether the calibration has any practical value. An INR is merely a guide to the prothrombin ratio that would have been obtained if the IRP [67/40] had been used. Such estimates are valueless without an indication of their precision.”

In the third substudy (chapter 5) we proposed a statistical model for INR estimates obtained in a randomly chosen laboratory, given the true INR is known. This model is based on our experiences in substudy 2, as well as on a number of reports on international calibration exercises. The model describes and quantifies central aspects of the INR/ISI system, such as the inaccuracy of INR estimates based on a given path of calibrations.

The model was used to derive an effective standard error of ISI estimates, which can be used for a rough calculation of the minimum CV of INR estimates (see page 76). In future calibration exercises it is recommended to
report the effective standard error instead of the usually reported marginal standard error.

In a simulation study it was demonstrated that the theoretically derived estimators had good statistical properties.

On page 83 the difference between the marginal and the effective standard error is demonstrated in a case where the correlation between calibrations within the same laboratory is zero. This also demonstrates the very important point that the effective standard error has two components: one, which may be reduced by increasing the number of participating laboratories, and another, which is due to inter-laboratory variation and which can only be reduced by perfecting the calibration methodology.

A test for the significance of the inter-laboratory variation was also provided. This test may be used to indicate the significance of the difference between the marginal and the effective standard error.

In an example, in which we investigated the first generation international reference preparation, RBT/79, we found that the 95% prediction interval for INR estimates, obtained directly by this IRP, is 1.7–2.4 INR and 3.2–5.0 INR for the true INR being 2 and 4, respectively. These intervals are large, compared with the therapeutic range of 2–3 INR for patients with mechanical heart valves. In routine use, the prediction interval is even larger, since routine INR estimates are obtained by commercial thromboplastins. The ISI of these thromboplastins are assigned either by the use of plasma calibrators, which have had their INR determined against an IRP, or by successive calibrations against a house standard and an IRP. Moreover, the imprecision of IRPs will generally increase with the distance from the primary IRP 67/40.

The results on inaccuracy can be seen as minimal estimates in the case, where the model is a good description of the reality. If the assumptions of linearity and unbiased estimation are not fulfilled, the inaccuracy will increase. Unfortunately, the WHO guidelines [102] are of very few words concerning the validation of these assumptions. Among the most important defects with respect to calibrations of IRPs are:

1. The desired inference space (time horizon and geography) for ISI estimates are not considered.

2. There is no definition of normal plasma other than it is “plasma obtained from a healthy person”. Moreover, no selection criteria or other directions are given on how to obtain normal plasma.

3. There are no recommendations on visual inspection or qualified statistical evaluation of the scatter plots of log PTs.
4. The crucial assumptions of log-linearity, equal size of the errors on the two axes, and, for the sake of the confidence limits, log-normal distribution, are not addressed.

5. There are no considerations whether the orthogonal regression line should be estimated from only abnormal plasma, or from all plasma samples on the same footing, or by a third method. Apparently, it is understood that the orthogonal regression is performed using all the plasma samples without distinguishing between normals and abnormalities.

6. No directions are given on the calculation of credibility intervals for individual INR estimates.

It is worrying that the guidelines recommend that samples with INRs outside the 1.5–4.5 range should be excluded. If the INR estimate is based on the reference thromboplastin, which is plotted on the vertical axis, the removal of points with large y-values will lead to systematic underestimation of the slope of the line.
Conclusion and perspectives

In the first substudy we found that the precision of patients’ own measurements of INR performed at home on CoaguChek coagulometers is sufficient for self management of OAT.

In the second substudy we found that the data were consistent with the hypotheses $H_1$ and $H_2$ for comparisons between all three methods. Thus, there is no reason to believe that the INR standard is invalid for any of the methods Nycotest and Hepato Quick. Moreover, there was no clinically significant deviation between INR measured by the two methods. The results indicate that it is possible to obtain accurate INR estimates by CoaguChek.

In the third substudy we demonstrated that it is possible to provide a good description of INR estimates, given a true value. Based on results from the statistical model it is recommended that in calibration exercises the *effective* standard error of the ISI is reported instead of the traditionally reported marginal standard error. The effective standard error contains sufficient information on the interlaboratory variance, such that it can be used directly for the calculation of the coefficient of variation of INR estimates. In an example, in which we examined the ISI of RBT/90, the effective standard error of the ISI was found to be minimum 0.070. In comparison the marginal standard error, which was reported, is only 0.027. Thus, the use of the marginal standard error may result in gross overestimation of the precision of INR estimates. This may explain some of the discrepant and “inexplicable” INR values, which are frequently reported in the scientific literature.

The all-important problem, which remains to be solved, is to improve the accuracy of INR estimates, while still retaining the definition of INR as the PT ratio that would have been obtained by using the first IRP 67/40 for PT determination. One way to proceed with this may be by collecting and analyzing all the data from the calibrations of IRPs in order to obtain an optimal description of each of the current IRP’s relation to the first IRP. It
is possible that a major improvement can be obtained simply by describing these relations as linear with two parameters.

To improve the description of the INR standard, the model in chapter 5 should be extended to cover cross-calibrations of IRPs. The model should also be used to calculate effective standard errors for all the IRPs that are currently available.

The influence of the assumption of equal variances in the calibration of thromboplastins should be closer investigated. The components of variance in the calibration should also be further described.

The problem, which initiated the present work, was that of establishing an optimal, computerized monitoring system for patients on self-managing anticoagulant therapy at Skejby Sygehus. This still remains to be implemented, but the solution has come much closer with the results presented in this thesis.
Appendix A

Some additional statistical results

A.1 Plots of differences between measurement methods

Suppose that some entity is measured by two different methods, A and B. Let \( X \) and \( Y \) denote the values obtained by A and B, respectively, and let \( Z \) denote the true value, which for simplicity is assumed to have a normal distribution with mean \( \mu \) and variance \( \tau^2 \). Moreover, assume that \( X \) and \( Y \) are unbiased estimates of \( Z \) and that the measurement errors have normal distributions with variances \( \sigma_x^2 \) and \( \sigma_y^2 \), respectively. Then \((X,Y)'\) has a twodimensional normal distribution, 

\[
\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \tau^2 + \sigma_x^2 & \tau^2 \\ \tau^2 & \tau^2 + \sigma_y^2 \end{pmatrix} \right\}.
\]

Now, the mean difference between \( Y \) and \( X \), given the unknown true value \( Z \), is

\[
E(Y - X \mid Z) = 0.
\]

However, the mean difference, given the measured value \( X \), is

\[
E(Y - X \mid X) = \left( \frac{\tau^2}{\tau^2 + \sigma_x^2} - 1 \right) X - \left( \frac{\tau^2}{\tau^2 + \sigma_x^2} - 1 \right) \mu.
\]

Note that the term within parentheses is strictly negative, unless \( \sigma_x^2 = 0 \), i.e. unless \( X \) is the true value.

Thus, if \( X \) is measured with error, a plot of \( Y - X \) against \( X \) will show a negative trend, even though both \( X \) and \( Y \) are unbiased estimates of the true value.
To see the relation between the regression lines of $Y$ on $X$ and of $Y - X$ on $X$, note that the mean of $Y$ given $X$ is

$$E(Y \mid X) = \frac{\tau^2}{\tau^2 + \sigma_x^2}X - \left(\frac{\tau^2}{\tau^2 + \sigma_x^2} - 1\right)\mu.$$ 

Thus the two lines have the same intercept, and the slope of the line relating $Y - X$ to $X$ is the slope of $Y$ against $X$ minus one. By this way, this property is independent of the assumption of normal distributions.

If the variances of the measurement errors are of the same magnitude, i.e. $\sigma^2_x \approx \sigma^2_y$, a plot of differences against sums or averages will be approximately without trend. To see this, note that the mean of $Y - X$ given $X + Y$ is

$$E(Y - X \mid X + Y) = \frac{\sigma_y^2 - \sigma_x^2}{\sigma_x^2 + \sigma_y^2 + 4\tau^2}(X + Y - 2\mu).$$

### A.2 The lognormal distribution

**Theorem 5** Let $X$ have the two-parameter lognormal distribution,

$$\log(X) \sim N(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \quad \sigma \in \mathbb{R}_+$$

Then the mean and variance of $X$ is

$$E\{X\} = \exp\left(\mu + \frac{1}{2}\sigma^2\right),$$

$$\text{Var}\{X\} = \exp\left(2\mu + \sigma^2\right)\left(e^{\sigma^2} - 1\right)$$

Moreover, for $\sigma^2 \ll 1$, the coefficient of variation of $X$ is

$$\text{CV}(X) = \left\{\exp(\sigma^2) - 1\right\}^{1/2} \approx \sigma$$

**Proof:** Let $Y = \log(X)$. Then, since $Y$ is normally distributed, the Laplace transform of $Y$ is

$$L_Y(t) = E\{e^{tY}\} = \exp\left(t\mu + \frac{1}{2}t^2\sigma^2\right)$$

Thus, the mean of $X$ is $L_Y(1)$ and the variance is $L_Y(2) - L_Y(1)^2$, and by insertion the desired result follows easily.

The CV of $X$ is given by

$$\text{CV}^2(X) = \frac{\text{Var}\{X\}}{E^2\{X\}} = \frac{\exp(2\mu + \sigma^2)\left(e^{\sigma^2} - 1\right)}{\exp(2\mu + \sigma^2)}$$

and the approximation follows from the fact that $\exp(x) \approx 1 + x$ for $|x| \ll 1.$
Appendix B

SAS programs

The following SAS macros and programs are available from the author.

B.1 Plotting and testing in calibrations

The following SAS macros were used to make the plots in figures 4.6-4.7.
APPENDIX B. SAS PROGRAMS

B.1.1 The CALIB macro

```sas
/* *************************************************************************************/
/* SAS MACRO */
/* */
/* NAME: CALIB */
/* */
/* DESCRIPTION: Performs errors-in-variables regression for INR with accompanying */
/* plots and tests. */
/* */
/* USED SAS MACROS: ereg.mac */
/* */
/* USED SAS MODULES: BASE, GRAPH. */
/* */
/* RESERVED NAMES: Names starting with an underscore (_), both */
/* variables and SAS datasets. */
/* */
/* OVERVIEW OVER VARIABLES IN MACRO CALL: */
/* data: Name of the SAS dataset containing the data. */
/* cat: Character variable with value 'p' or 'n', indicating */
/* patient (abnormal) and normal, respectively. */
/* x: Variable containing log PTs for the x-axis. */
/* y: Do. for the y-axis. */
/* delta: The ratio between measurement errors on the x- and */
/* y-axis. Is commonly assumed to be 1, although this is */
/* not always appropriate. */
/* */
/* EXAMPLE OF MACRO CALL: %calib(data=pt, cat=type, x=nt, y=crm, delta=1) */
/* */
/* AUTHOR: Joern Attermann. */
/* */
/* LAST UPDATED: 12-09-99 */
/* */
/* The macro is freeware, which means that it may be freely */
/* used. Moreover it may be freely distributed as long as no changes */
/* are made to the macro. */
/* */
/* NO LIABILITY: The author can accept no responsibility for loss or */
/* damage, whether real or consequential, arising from the use of the */
/* macro. By using the macro you explicitly agree to hold the author */
/* blameless for any such losses or damages. */
/* */
/*********************************************************/
%macro calib(data=, cat=, x=, y=, delta=1);

title "Errors-in-variables regression: &y vs. &x";

data _data;
set &data end=last;
retain _xmin _ymin 100000000 _xmax _ymax -100000000;
if &x=. & &y=. & length(&cat)>0 then do;
   _xmin = min(_xmin, &x);
   _xmax = max(_xmax, &x);
   _ymin = min(_ymin, &y);
   _ymax = max(_ymax, &y);
output;
end;
if last then do;
```
B.1. PLOTTING AND TESTING IN CALIBRATIONS

call symput('xmin', _xmin);
call symput('xmax', _xmax);
call symput('ymin', _ymin);
call symput('ymax', _ymax);
end;
run;

title2 'Abnormals';
data _pts;
set _data;
if upcase(&cat) = 'P';
run;
%ereg(data=_pts, y=&y, x=&x, delta=&delta);
data _null_;
set _ereg end=last;
if last then do;
call symput('icept', _b0hat); /* Intercept estimate */
call symput('slope', _b1hat); /* Slope estimate */
call symput('v_icept', _v_b0hat); /* Variance of intercept estimate */
call symput('v_slope', _v_b1hat); /* Variance of slope estimate */
call symput('cov_slic', _v_b0b1); /* Covariance of slope and intercept estimates */
call symput('v_u', _su_hat); /* Variance of measurement error of x */
call symput('m_x', _m_x); /* Mean of X */
end;
run;

title2 'Normals';
data _normals;
set _data end=last;
retain _nnormal 0;
call symput('nnormal', '0');
if upcase(&cat) = 'N' then do;
    _nnormal = _nnormal + 1;
    output;
end;
if last then call symput('nnormal', _nnormal);
run;
%if &nnormal>0 %then %do;
%ereg(data=_normals, y=&y, x=&x, delta=&delta);
data _null_;
set _ereg end=last;
if last then do;
call symput('m_x', _m_x); /* Mean of normal X */
call symput('m_y', _m_y); /* Mean of normal Y */
call symput('v_mx', _v_mx); /* Variance of mean normal X */
call symput('v_my', _v_my); /* Variance of mean normal Y */
call symput('cmxmy', _v_mxmy); /* Cov of mean normal X and Y */
end;
run;

title2 'Test'; /* Test for Mean Normal log PT on errors-in-variables regression */
/* line for abnormal plasma: */
data _null_; file print;
dist = &mny - &icept - &slope * &mnx;
/* Var(dist) according to Tomenson: */
v_dist = &vmny + &slope**2*&vmnx - 2*&slope*&cmnxmny
 + &v_icept + 2* &vmnx* &cov_slic + &mnx**2* &v_slope;
z = dist / sqrt(v_dist);
put;
p = 2*(1-probnorm(abs(z)));
put 'Test for Mean Normal log PT on errors-in-variables
regression line for abnormal plasma: p = ' p;
run;
%end;
%else %do;
%let mnx = .;
%let mny = .;
%let vmnx = .;
%let vmny = .;
%let cmnxmny = .;
%end;

data _lines;
/* Errors-in-variables regression line with 95% confidence limits: */
do &x = &xmin to &xmax by (&xmax-&xmin)/500;
_y = &icept + &slope * &x; /* Estimate */
/* sd for regression line: */
_sdvy = sqrt(&v_icept + &x**2 * &v_slope + 2 * &x * &cov_slic);
/* lower limit for regression line: */
_ylo = _y - 1.96 * _sdvy;
/* upper limit for regression line: */
_yhi = _y + 1.96 * _sdvy;
/* sd for individual observations of Y given X observed: */
_sdindy = sqrt((1-(&slope**2 * &v_x**2)/(&slope**2* &v_x + &delta* &v_u)/(&v_x + &v_u))
 + (&slope**2 * &v_x + &delta* &v_u));
/* mean of individual observation of Y given X observed: */
_mindy = &icept + &slope* &m_x + &slope* &v_x/(&v_x + &v_u)*(&x - &m_x);
_yindlo = _mindy - 1.96_*sdindy;
_yindhi = _mindy + 1.96_*sdindy;
&cat = 'n'; /* Necessary to produce plots */
output;
end;
%if &nnormal>0 %then %do;

/* (1-alpha) confidence ellipse for mean normal log PT: */
_alpha = 0.05;
_t2 = (%nnormal-1)*2/(%nnormal-2)*finv(1-_alpha,2,%nnormal-1);
_mx = &mnx;
_my = &mny;
_sx = sqrt(&vmnx);
_sy = sqrt(&vmny);
_r = &cmnxmny / _sx / _sy;
_xlo = &mnx - sqrt(_t2 * &vmnx);
_xhi = &mnx + sqrt(_t2 * &vmnx);
_y = .;
_ylo = .;
B.1. PLOTTING AND TESTING IN CALIBRATIONS

\begin{verbatim}
_yhi = .;
_yindhi = .;
_yindlo = .;
_ylo = .;
_yhi = .;

_do &x=_xlo to _xhi by (_xhi - _xlo)/500;
   _determ = ( &cmnxmny**2 - &vmnx* &vmny )/ &vmnx * ( ( &x - &mnx )**2 / &vmnx - _t2 );
   _ynhi = &mny + &cmnxmny / &vmnx * ( &x - &mnx ) + sqrt(_determ);
   _ynlo = &mny + &cmnxmny / &vmnx * ( &x - &mnx ) - sqrt(_determ);
   &cat = 'n'; /* Necessary to produce plots */
   output;
end;

/* Mean normal log PT: */
_y = .;
_ylo = .;
_yhi = .;
_yindhi = .;
_yindlo = .;

&_y = &mny;
&_x = &mnx;
&_cat = 'n';
output;

run;

data _total;
   set _pts _lines;
run;

proc sort data=_total;
   by &x;
run;

title2;
goptions reset=symbol;
axis1 order=( &ymin to &ymax by 0.1 );
axis2 order=( &xmin to &xmax by 0.1 );
   * axis1 order=(2 to 5 by 0.1 );
   * axis2 order=(2 to 5 by 0.1 );
   symbol1 i=none v=plus h=2;
   symbol2 i=none v=circle h=0.4;
   symbol3 i=j l=2;
   symbol4 i=j;
   symbol5 i=j l=2;
   symbol6 i=j l=4;
   symbol7 i=j l=4;
   symbol8 i=j l=2;
   symbol9 i=j l=2;
proc gplot data=_total;
   plot &y* &x=&cat / vaxis=axis1 haxis=axis2;
   plot2 ( _ylo _y _yhi _yindlo _yindhi _ynlo _ynhi )* &x / overlay vaxis=axis1 haxis=axis2;
run;
quit;
%mend calib;
\end{verbatim}
Example of output from the CALIB macro

The macro call

```sas
%calib(data=standard, y=crm1, x=hq, cat=cat, delta=1);
```

produces the following output:

```
Errors-in-variables regression: crm1 vs. hq
Abnormals
**********************************************************************************
Slope estimate: 0.7541719677 with standard deviation 0.030529351
and variance 0.0009255277
Intercept estimate: 0.3983083365 with standard deviation 0.120441835
and variance 0.0145068013
Correlation of slope and intercept estimates: -0.997675348 ; covariance: -0.003675448
Standard deviation of true x: 0.2848048919 ; variance: 0.081158771
Standard deviation of true y: 0.2149465808 ; variance: 0.0461367119
Standard deviation of measurement error of X: 0.049679791 ; variance: 0.0024680816
Standard deviation of measurement error of Y: 0.0487945808 ; variance: 0.0024680816
Estimate of mean of X: 3.9270642962 with standard deviation 0.0393427387
and variance 0.0015368013
Estimate of mean of Y: 3.3599901442 with standard deviation 0.0300014795
and variance 0.0009000888
Estimate of correlation between estimates of mean of X and mean of Y: 0.9597878206 ;
covariance: 0.0011328763
**********************************************************************************
Slope estimate: 0.7381552176 with standard deviation 0.2728928828
and variance 0.07447058
Intercept estimate: 0.4902497752 with standard deviation 0.8455976274
and variance 0.7150353475
Correlation of slope and intercept estimates: -0.999831983 ;
covariance: -0.230718888
Standard deviation of true x: 0.0660616496 ; variance: 0.00436514061
Standard deviation of true y: 0.0487637513 ; variance: 0.0023779034
Standard deviation of measurement error of X: 0.0371804233 ; variance: 0.0013823839
Standard deviation of measurement error of Y: 0.0371804233 ; variance: 0.0013823839
Estimate of mean of X: 3.0981212642 with standard deviation 0.0239719115
and variance 0.0005746525
Estimate of mean of Y: 2.7714141511 with standard deviation 0.0193914603
and variance 0.0003760287
Estimate of correlation between estimates of mean of X and mean of Y: 0.6930001685 ;
covariance: 0.0003221414

Errors-in-variables regression: crm1 vs. hq
Test

Test for Mean Normal log PT on orthogonal regression line for abnormal plasma: p = 0.1659455326
```
B.1. PLOTTING AND TESTING IN CALIBRATIONS

B.1.2 The EREG macro

```sas
/* ******************************************************
* NAME: EREG
* DESCRIPTION: Performs errors-in-variables regression according to
* USED SAS MODULES: BASE.
* RESERVED NAMES: Names starting with an underscore (_), both
* variables and SAS datasets.
* OVERVIEW OVER VARIABLES IN MACRO CALL:
* data: Name of the SAS dataset containing the data.
* x: Variable for the x-axis.
* y: Do. for the y-axis.
* delta: The ratio between variances of measurement errors on
* the y- and x-axis.
* diags: Require dataset for diagnostic plots? ('yes' or 'no')
* NOTE: Missing data are not allowed. Validity of variance estimates
* requires twodimensional normal distribution of X and Y.
* If diagnostics are required (by setting diags = 'yes' in the macro
* call), a dataset _diags will contain best linear unbiased
* estimates (BLUEs) of x ( _x_hat ) and residual ( _v_hat).
* EXAMPLE OF MACRO CALL: %ereg(data=pt, x=nt, y=crm, delta=1, diags=no)
* AUTHOR: Joern Attermann.
* LAST UPDATED: 10-09-99
* The macro is freeware, which means that it may be freely
* used. Moreover it may be freely distributed as long as no changes
* are made to the macro.
* NO LIABILITY: The author can accept no responsibility for loss or
* damage, whether real or consequential, arising from the use of
* the macro. By using the macro you explicitly agree to hold the
* author blameless for any such losses or damages.
* *******************************************************/

sum of observed X
sum of squares of observed X
sum of observed Y
sum of squares of observed Y
sum of products of observed X and Y
mean of observed X
mean of observed Y
empirical variance of observed X
empirical variance of observed Y
empirical covariance of observed X and Y
```
APPENDIX B. SAS PROGRAMS

/* _b1hat estimate of slope of orthogonal regression line */
/* _sx_hat estimate of variance of true x */
/* _su_hat estimate of variance of measurement error on X */
/* _se_hat estimate of variance of true y */
/* _svv estimate of variance of measurement error on Y */
/* _sxv estimate of variance of slope estimate */
/* _v_b0hat estimate of variance of intercept estimate */
/* _v_b0b1 estimate of covariance of intercept and slope estimates */
/* _v_mx estimate of variance of mean of X */
/* _v_my estimate of variance of mean of Y */
/* _v_mxmy estimate of covariance of mean of X and mean of Y */

%macro ereg(data=, x=, y=, delta=1, diags=no);
data _ereg;
    set &data end=last;
    retain _sx _sx2 _sy _sy2 _sxy 0;
    _sx = _sx + &x;
    _sx2 = _sx2 + (&x)**2;
    _sy = _sy + &y;
    _sy2 = _sy2 + (&y)**2;
    _sxy = _sxy + (&x)*(&y);
    if last then do;
        _m_x = _sx/_n_;  
        _m_y = _sy/_n_;  
        _mxx = (_sx2 - _sx**2/_n_) / (_n_-1);  
        _myy = (_sy2 - _sy**2/_n_) / (_n_-1);  
        _mxy = (_sxy - _sx*_sy/_n_) / (_n_-1);  
        _b1hat = (_myy - &delta*_mxx + sqrt((_myy-&delta*_mxx)**2 + 4*&delta*_mxy**2))/2/_mxy;  
        _sx_hat = 1/(2*&delta)* (_myy + &delta*_mxx - sqrt((_myy-&delta*_mxx)**2 + 4*&delta*_mxy**2));  
        _su_hat = 1/(2*&delta)*(_myy - &delta*_mxx + sqrt((_myy-&delta*_mxx)**2 + 4*&delta*_mxy**2));  
        _suvhat = -_b1hat*_su_hat;
        _svv = (_n_-1)/(_n_-2)*(&delta+_b1hat**2)*_su_hat;
        _b0hat = _m_y - _b1hat * _m_x;
        _v_b1hat = (_sx_hat*_svv + _su_hat*_svv - _suvhat**2)/(_n_-1) / _sx_hat**2;
        _v_b0hat = _svv/_n_ + _m_x**2 + _v_b1hat;
        _v_b0b1 = _m_x * _v_b1hat;
        _sy_hat = _b1hat**2 * _sx_hat;
        _se_hat = &delta * _su_hat;
        _v_mx = (_sx_hat + _su_hat) / _n_;  
        _v_my = (_b1hat**2*_sx_hat + _su_hat) / _n_;  
        _v_mxmy = _b0hat * _sx_hat;
    %if %upcase(&diags) = YES %then %do;
        call symput('_b1hat', _b1hat);
        call symput('_b0hat', _b0hat);
        call symput('_su_hat', _su_hat);
        call symput('_se_hat', _se_hat);
    %end;
    file print;
    put '*****************************************************************************';
    dum = sqrt(_v_b1hat);
    put 'Slope estimate: ' _b1hat ' with standard deviation ' dum ' and variance ' _v_b1hat;
    dum = sqrt(_v_b0hat);
    put 'Intercept estimate: ' _b0hat ' with standard deviation ' dum ' and variance ' _v_b0hat;
    dum = _v_b0b1 / sqrt(_v_b1hat) / sqrt(_v_b0hat);
%mend;
B.1. PLOTTING AND TESTING IN CALIBRATIONS

put 'Correlation of slope and intercept estimates: ' dum ' ; covariance: ' _v_b0b1;
dum = sqrt(_sx_hat);
put 'Standard deviation of true x: ' dum ' ; variance: ' _sx_hat;
dum = sqrt(_sy_hat);
put 'Standard deviation of true y: ' dum ' ; variance: ' _sy_hat;
dum = sqrt(_su_hat);
put 'Standard deviation of measurement error of X: ' dum ' ;
  variance: ' _su_hat;
dum = sqrt(_se_hat);
put 'Standard deviation of measurement error of Y: ' dum ' ;
  variance: ' _se_hat;
dum = sqrt(_v_mx);
put 'Estimate of mean of X: ' _m_x ' with standard deviation ' dum ';
  variance: ' _v_mx;
dum = sqrt(_v_my);
put 'Estimate of mean of Y: ' _m_y ' with standard deviation ' dum ';
  variance: ' _v_my;
dum = _v_mxmy / sqrt(_v_mx) / sqrt(_v_my);
end;
run;
%if %upcase(&diags) = YES %then %do;
data _diags;
  set &data;
  /* BLUE for true x: */
  _x_hat = (&_su_hat * (&y - &_b0hat) * &_b1hat + &_se_hat * &x)
            / (&_b1hat**2 * &_su_hat + &_se_hat);
  /* BLUE for residual: */
  _v_hat = _y - &_b0hat - &_b1hat* &_x;
run;
%end;
%end ereg;

Example of output from the EREG macro

The macro call
%ereg(data=temp, x=hq, y=crm1, diags=yes);

produces the following output:

The SAS System
******************************************************************************
Slope estimate: 0.7541719677 with standard deviation 0.0305929351
and variance 0.0009359277
Intercept estimate: 0.3983083365 with standard deviation 0.1204441835
and variance 0.0145068013
Correlation of slope and intercept estimates: -0.997477998 ; covariance: -0.003675448
Standard deviation of true x: 0.2848084919 ; variance: 0.081158771
Standard deviation of true y: 0.2147945808 ; variance: 0.0461367119
Standard deviation of measurement error of X: 0.049679791 ; variance: 0.0024680816
Standard deviation of measurement error of Y: 0.049679791 ; variance: 0.0024680816
Estimate of mean of X: 3.9270642962 with standard deviation 0.0393427387
and variance 0.0015478511
Estimate of mean of Y: 3.35999901442 with standard deviation 0.0300014795
and variance 0.0009000888
Estimate of correlation between estimates of mean of X and mean of Y: 0.9597878206 ;
covariance: 0.0011328763
B.2 Estimating interlaboratory variance

This program was used for estimating and testing interlaboratory variance in section 5.5

title 'Calibration of RBT/79 against IRP 67/40';

data cal;
input i x s n;
nu = n-2;
cards;
1 1.482 0.036 76
2 1.317 0.036 72
3 1.591 0.028 79
4 1.374 0.047 47
5 1.33 0.041 44
6 1.374 0.041 69
7 1.423 0.037 77
;
run;

/************************************************************/
/* First the full model */
/************************************************************/

data fuld;
set cal end=last;
k = _n_;array ax{50} ax1-ax50;array as{50} as1-as50;array anu{50} anu1-anu50;array a{50};array b{50};array gamma{50};array tau{50};array z{3} z1-z3;array sig2i{50};array sigi{50};retain ax1=ax50 as1=as50 anu1=anu50;ax{i} = x;as{i} = s;anu{i} = nu;
pi = arcos(-1);
if last then do;
   /* Find start estimates */
   s_x = 0; /* Beregn mu */ms_s = 0; /* og ms_s */do i=1 to k;
   s_x = s_x + ax{i};ms_s = ms_s + as{i}**2;
end;
mu = s_x / k;ms_s = ms_s / k;
msd_x = 0; /* Beregn MSD_x */
do i=1 to k;
    msd_x = msd_x + (ax{i} - mu)**2;
end;
msd_x = msd_x / (k-1);
sigma2 = msd_x - ms_s;
/* Iterer */
do iter=1 to 1000;
do i=1 to k;
    a{i} = sigma2 / (ax{i} - mu)**2;
    b{i} = as{i}**2 / (anu{i}+1) / (ax{i} - mu)**2;
/* Find solutions to the equation */
/* gamma^3 - (a+2)*gamma^2 + {(nu+2)*a + nu*b + 1}*gamma - (nu+1)*a = 0 */
coeff2 = -(a{i} + 2);
coeff1 = (anu{i}+2)*a{i} + anu{i}*b{i} + 1;
coeff0 = -(anu{i}+1)*a{i};
p = (3*coeff1-coeff2**2)/3;
q = (9*coeff1*coeff2 - 27*coeff0 - 2*coeff2**3)/27;
QQ = p/3;
RR = q/2;
DD = QQ**3 + RR**2;
/* Test for 1 or 3 roots in [0,1] (Vangel, appendix): */
qi = -2 -6*sqrt(anu{i}+1) * sin( (arsin( sqrt( anu{i}/(anu{i}+1) ) ) -pi/2 )/3 );
hi = (1-qi)**3 / 27 / anu{i};
if (ax{i} - mu)**2 < max(sigma2/qi, as{i}**2 / (anu{i}+1) / hi ) then do;
    /* Only one root: */
    if DD >= 0 then do;
        if RR = sqrt/DD >= 0 then SS = (RR + sqrt/DD)**(1/3);
        else SS = -(RR + sqrt/DD)**(1/3);
        TT = (RR - sqrt/DD)**(1/3);
        z{1} = -coeff2/3 + SS + TT;
        z{2} = .,
        z{3} = .;
        roots = 1;
gamma{i} = z{1};
    end;
    else do;
        theta = arcos( RR / sqrt(-QQ**3) );
        z{1} = 2*sqrt(-QQ) * cos(theta/3) - coeff2/3;
        z{2} = 2*sqrt(-QQ) * cos( (theta+2*pi)/3 ) - coeff2/3;
        z{3} = 2*sqrt(-QQ) * cos( (theta+4*pi)/3 ) - coeff2/3;
        gamma{i} = .;
        if z{1}>0 & z{1}<1 then gamma{i} = z{1};
        if z{2}>0 & z{2}<1 then gamma{i} = z{2};
        if z{3}>0 & z{3}<1 then gamma{i} = z{3};
        if gamma{i}=., then do;
            put 'Iteration ' iter ', i = ' i ',GAMMAi = .';
            gamma{i} = 0.5;
        end;
        end;
    end;
else do;
    /* 3 roots: */
B.2. ESTIMATING INTERLABORATORY VARIANCE

put 'ALARM!!!';
roots = 3;
end;

test1 = z1**3 + coeff2*z1*z2 + coeff1*z1 + coeff0;
test2 = z2**3 + coeff2*z2*z3 + coeff1*z2 + coeff0;
test3 = z3**3 + coeff2*z3*z2 + coeff1*z3 + coeff0;
end;
/* Update sigma2 and mu */
staeller = 0;
snaevner = 0;
mtaeller = 0;
mnaevner = 0;
do i=1 to k;
    staeller = staeller + gamma{i} * ((ax{i}-mu)**2 + anu{i}*as{i}**2/(anu{i}+1)/(1-gamma{i}));
    snaevner = snaevner + anu{i} + 1;
    mtaeller = mtaeller + gamma{i}*ax{i};
    mnaevner = mnaevner + gamma{i};
end;
sigma2 = staeller/snaevner;
mu = mtaeller/mnaevner;
/* Calculate -2*loglikelihoodfunction */
like=0;
do i=1 to k;
    like = like - (anu{i}+1)*log(gamma{i}/sigma2) + gamma{i}/sigma2*((ax{i}-mu)**2 + anu{i}*as{i}**2/(anu{i}+1)/(1-gamma{i}))) + anu{i}*log(1-gamma{i}) + log(2*pi)*(anu{i}+1) + (anu{i}+1)*log(anu{i}+1);
end;
output;
file print;
put '*******************************************************';
put 'Full model';
put 'mu = ' mu;
put 'sigma2 = ' sigma2;
put '-2loglike = ' like;
do i=1 to k;
    sig2i{i} = (anu{i}+1) * sigma2 * (1-gamma{i}) / gamma{i};
    sigi{i} = sqrt(sig2i{i});
    put 'sig2i (i=' i ') = ' sig2i{i} ' sigi (i=' i ') = ' sigi{i};
end;
format mu 12.6 sigma2 12.6 like 12.6;
run;

/*****************************/
/* Now the model with sigma2 = 0 */
/*****************************/
data redukt;
  set cal end=last;
  k = _n_;  
  array ax{50} ax1-ax50;
  array as{50} as1-as50;
  array anu{50} anu1-anu50;
  array tau{50};
  array sig2i{50};
  array sigi{50};
  retain ax1-ax50 as1-as50 anu1-anu50;
  ax{i} = x;
  as{i} = s;
  anu{i} = nu;
  pi = arcos(-1);
  if last then do;
  /* Find start estimates */
    s_x = 0;  /* Calculate mu */
    do i=1 to k;
      s_x = s_x + ax{i};
    end;
    mu = s_x / k;
  /* Iterate! */
  do iter=1 to 1000;
  /* Calculate tau_i**2: */
    do i=1 to k;
      tau{i} = ( (ax{i}-mu)**2 + anu{i}*as{i}**2/(anu{i}+1) ) / (anu{i}+1);
    end;
  /* Calculate mu */
    mtaeller = 0;
    mnaevner = 0;
    do i=1 to k;
      mtaeller = mtaeller + ax{i}/tau{i};
      mnaevner = mnaevner + 1/tau{i};
    end;
    mu = mtaeller / mnaevner;
  /* Calculate -2loglike except for a constant: */
  like = 0;
  do i=1 to k;
    like = like + (anu{i}+1) * log(tau{i})
         + 1/tau{i} * ( (ax{i}-mu)**2 + anu{i}*as{i}**2 / (anu{i}+1) )
         + log(2*pi)*(anu{i}+1) * log(anu{i}+1);
  end;
  output;
end;
file print;
put;
put '*****************************************************';
put;
put 'Reduced model';
put;
put 'mu = ' mu;
put ' -2loglike = ' like;
  do i=1 to k;
    sig2i{i} = tau{i} * (anu{i}+1);
    sigi{i} = sqrt(sig2i{i});
    put 'sig2i (i=' i ') = ' sig2i{i} ' sigi (i=' i ') = ' sigi{i};
end;
end;
run;
B.3 Simulation of $\log(\log(\text{INR}))$

The following SAS program was used to perform the simulations in section 5.3.1.

```sas
proc options (nullvalue = .)
/
* Simulation study of the model for $\log(\log(\text{INR}))$. */
* JA 08-11-99 */
* Models 3 steps ($r=3$) in fictive calibration exercise. */
* Assumes no within-lab correlation ($\rho$) between calibrations. */
* Uses empirical result: intralab log(sd) of slope estimate */
* ( $\log(\hat{\nu}_{il})$ ) is approximately $\sim N(-4, 0.25^2)$. */
* True log(INR), $y$, is sampled from a $N(1, 0.4^2)$ distribution, */
* which is truncated at $\log(1.2)$ and $\log(8)$. */

data siminr;

/* Parameters: */
nsim = 1000;
point = 0;  /* Enumeration of points in the grid */
do b1 = 0.8 to 1.2 by 0.2; /* True slope in first calibration */
do b2 = 0.8 to 1.2 by 0.2; /* do. second calibration */
do b3 = 0.8 to 1.2 by 0.2; /* do. third calibration */
do gamma1 = 0 to 0.1 by 0.05; /* Interlab sd in first calibration */
do gamma2 = 0 to 0.1 by 0.05; /* do. second calibration */
do gamma3 = 0 to 0.1 by 0.05; /* do. third calibration */
do until(INR>1.2 & INR<8); /* Sample the true INR */
  y = 1 + 0.4 * normal(0);
  INR = exp(y);
  k1 = 7;  /* Number of labs in first calibration */
  k2 = 16;  /* do. second calibration */
  k3 = 11;  /* do. third calibration */
  tau = 0.05;  /* sd between normal persons */
  sigma = 0.04;  /* Analytical CV of PT measurements */
  mu = 3;  /* Mean normal log PT */
  n = 20;  /* Number of normal persons */
array b (3) b1-b3;
array gamma (3) gamma1-gamma3;
array k (3) k1-k3;
array kappa (3);
array zeta (3);
array nu (3);  /* Square root of the average of the true sd's of the */
  /* slope estimates in i'th calibration. */
c = b1*b2*b3;  /* True common ISI */
end;
```

The following SAS program was used to perform the simulations in section 5.3.1.
B.3. SIMULATION OF LOG(LOG(INR))

nytil = 0;
sytil = 0;
mlnytil = 0;
slytil = 0;
mreldif = 0;
sreldif = 0;

/* Simulation: */
do sim=1 to nsim;
  /* sim'th simulation: */
c_til = 1; /* Common ISI-estimate */
do i=1 to 3;
    /* i'th calibration step: */
b_i_til = 0; /* Common slope estimate in i'th calibration step */
  nu_dum = 0;
do l=1 to k{i}; /* l'th laboratory: */
b_l = b{i} + gamma{i}*normal(0); /* True slope in l'th lab */
  nu_il = exp(-4 + 0.25*normal(0)); /* True sd of slope estimate in l'th lab */
  nu_dum = nu_dum + nu_il**2;
  b_i_til = b_i_til + b_l + nu_il*normal(0); /* Slope estimate in l'th lab */
  b_i_til = b_i_til / k{i};
  b_i = b_i_til / k{i};
  nu{i} = sqrt( nu_dum / k{i} );
c_til = c_til * b_i_til;
end;
/* INR determination in a specific lab, lambda: */
c_lambda = 1; /* True ISI of lambda */
do i=1 to 3;
    /* i'th calibration step: */
b_i_lam = b{i} + gamma{i}*normal(0); /* True slope in lambda */
c_lambda = c_lambda * b_i_lam;
end;
u = mu + sqrt( (tau**2+sigma**2)/n ) * normal(0); /* Estimated mean log PT */
x = y/c_lambda + mu; /* True log(PT) of the given patient */
x_til = x + sigma*normal(0); /* Estimated log(INR) of the given patient */
y_til = mu_til + y_til; /* Empirical mean of log(INR) */
sytil = sytil + y_til**2; /* Empirical sd of log(INR) */
lly_til = log(y_til); /* Estimated log(log(INR)) */
if sim=nsim then do;
    mytil = mytil/nsim;
sytil = sqrt( (sytil - nsim*mytil**2)/(nsim-1) );
mlnytil = mlytil/nsim;
slytil = sqrt( (slytil - nsim*mlytil**2)/(nsim-1) );
sreldif = mreldif/nsim;
sreldif = sqrt( (sreldif - nsim*mreldif**2)/(nsim-1) );
end;

/* Calculate theoretical values from known values: */
i=1;
kappa{i} = gamma{i}/b{i};
  zeta{i} = sqrt( gamma{i}**2 / b{i}**2 / k{i} + 1/b{i}**2 /k{i} + nu{i}**2 );
do i=2 to 3;
    kappa{i} = sqrt( kappa{i-1}**2 + gamma{i}**2/b{i}**2 );
    zeta{i} = sqrt( zeta{i-1}**2 + gamma{i}**2 / b{i}**2 / k{i} + 1/b{i}**2 /k{i} + nu{i}**2 );
end;
omega = c_til * sqrt( sigma**2 + tau**2/n );
m_ly_til = log(y); /* Theoretical value for mean of log(log(INR)) */
\texttt{\textbackslash vartheta = sqrt( kappa(3)**2 + zeta(3)**2 + omega**2/y**2 );}
\texttt{CVINR = y * vartheta; /* Theoretical value for CV of INR */}
\texttt{/* Theoretical value for effective se of ISI: */}
\texttt{effISI = c * sqrt(zeta(3)**2 + kappa(3)**2);}
\texttt{output; end;}
\texttt{point = point + 1;}
\texttt{put 'Point in grid: ' point;}
\texttt{end; end; end; end; end; end; end; run;}

\texttt{/* For subsequent data analysis: */}
\texttt{data temp;}
\texttt{set siminr;}
\texttt{Ve = slytil;}
\texttt{Va = vartheta;}
\texttt{lve = log(ve);}
\texttt{lva = log(va);}
\texttt{vdiff = ve-va;}
\texttt{lvdiff = lve - lva;}
\texttt{Ue = mlytil;}
\texttt{Ua = m_ly_til;}
\texttt{udiff = ue-ua;}
\texttt{run;
Summary

Life-long oral anticoagulant therapy (OAT) with vitamin K antagonists is offered to patients with increased risk of thrombosis, e.g. patients with artificial heart valves or with atrial fibrillation. It is estimated that in 1992 in the Nordic countries 0.3 – 0.5% of the population was undergoing daily anticoagulant therapy.

The therapy necessitates close monitoring of coagulant activity, since excess doses of anticoagulant medicine may lead to life-threatening bleedings. Traditionally, patients on OAT are required to pay regular visits to a physician, who decides on drug dosage adjustments and the time for the next visit based on laboratory analyses of the INR. This conventional treatment regimen is relatively inconvenient for the patient, since it requires frequent outpatient visits and venipunctures. Moreover, errors may occur due to insufficient communication between patient and physician. There is general agreement that the quality of the therapy is too low, and often unexpected fluctuations in the coagulant activity are seen.

Recently, OAT based on patient self-management has become a realistic alternative by the availability of small portable whole blood coagulometers. An important part of the new concept is the training and continuous support and monitoring of the patients, and a center with these purposes has been established at Skejby Sygehus.

The main instrument for monitoring the coagulant activity is the prothrombin time (PT). This is the time until clotting can be observed in a plasma sample, in which the extrinsic coagulation cascade is triggered by the addition of a thromboplastin preparation. Thromboplastin is manufactured from human, rabbit, or bovine tissue, and the active properties vary significantly from species to species as well as from batch to batch within the same species, even though standardized methods of preparation are used.

In order to standardize the PT, the World Health Organization (WHO) recommended in 1983 that the International Normalized Ratio (INR) should be reported instead of the raw PT. The INR of a plasma sample is defined as the PT ratio for that sample to normal samples that would have been
obtained, had the International Reference Preparation (IRP) of thromboplastin, IRP 67/40, with a certain standard procedure, been used for the PT determination. The IRP 67/40 has, however, been out of supply for many years, and the present methodology for estimating INR is based on properties of IRPs, which have been calibrated against the IRP 67/40 or one of its successors.

Obviously, the quality of INR determinations is highly dependent on the properties of these secondary IRPs and the applied technique as well as on the quality of the estimates of these properties, which are obtained in the calibration exercises. Thus, in order to obtain reliable INR estimates for daily use in OAT and for establishing realistic dose-response relationships for OAT, it is imperative that the physical relations between successive IRPs are thoroughly investigated and described. Numerous reports on inconsistent or discrepant INR values suggest that the distributional properties of the INR system are not fully understood.

The aim of this Ph.D. thesis was to investigate and describe the INR system in order to provide a basis for the evaluation and comparison of different forms of anticoagulant therapy.

The specific hypotheses were:

- The precision of patient’s own measurements of INR performed at home on a portable coagulometer is sufficient to allow for self management of OAT (substudy 1).

- For selected pairs of thromboplastins, the relation between logarithmic prothrombin times is linear, such that the standardization by the INR system is feasible (substudy 2).

- Based on a realistic statistical model, the statistical distribution of INR estimates, given a true INR value, can be adequately described to allow for good estimates of accuracy and precision (substudy 3).

In the first substudy it was shown that for selected patients the precision of the patients’ own measurements of INR is sufficient to allow for reliable routine patient self testing of INR. In the same substudy we found large discrepancies between the INR measurements on portable coagulometers and in the Department of Clinical Chemistry, Skejby Sygehus. However, when available estimates of laboratory and CoaguChek variation were taken into consideration, the observed differences were not unlikely. In the same study we found that the deviation between the portable coagulometers and the Department of Clinical Chemistry, Roskilde County Hospital, was not larger than the deviation between the two clinical chemistry departments. The
material in our study is, however, not strong enough to draw any firm conclusions regarding the accuracy of neither portable coagulometers nor clinical chemistry departments.

In the second substudy we investigated the fundamental assumptions of the INR system. We found that the data from the comparison of three thromboplastin preparations (CRM 149S, Nycotest and Hepato Quick) were consistent with these assumptions and concluded that the INR system is valid for these thromboplastins.

In the third substudy we proposed a statistical model for INR estimates obtained in a randomly chosen laboratory, given the true INR is known. This model was based on our experiences in substudy 2, as well as on a number of reports on international calibration exercises. The model describes and quantifies central aspects of the INR system, such as the inaccuracy of INR estimates based on a given path of calibrations.

The main result states that, under weak regularity conditions, log (log (estimated INR)) is approximately normally distributed with mean log (log (true INR)). The variance is a function of the interlaboratory variance, the accuracy of the slope estimates in the calibrating laboratories, the accuracy of local PT determinations, and the interindividual variance of the normal persons used for normalizing the PT. A simulation study showed that the derived theoretical results are in close agreement with the empirical ones.

The third part of the thesis also contains a formula and test for the interlaboratory variation of slope estimates in a calibration exercise.

In an example, in which we investigated the first generation international reference preparation, RBT/79, we found that the 95% prediction interval for INR estimates, obtained directly by this IRP, is (1.7, 2.4) and (3.2, 5.0) for the true INR being 2 and 4, respectively. These intervals are large, compared with the therapeutic range of (2, 3) for patients with mechanical heart valves. In routine use, the prediction interval is even larger, since routine INR estimates are obtained by commercial thromboplastins. The ISI of these thromboplastins are assigned either by the use of plasma calibrators, which have had their INR determined against an IRP, or by successive calibrations against a house standard and an IRP. Moreover, the imprecision of IRPs will generally increase with the distance from the primary IRP 67/40.

Based on results from the statistical model it is recommended that in calibration exercises an effective standard error of the ISI is reported instead of the traditionally reported marginal standard error. The effective standard error contains sufficient information on the interlaboratory variance, such that it can be used directly for the calculation of the coefficient of variation of INR estimates. In the example given above, the effective standard error of the ISI is 0.0956, and in comparison the marginal standard error is only
0.0364. Thus, the use of the marginal standard error will result in gross overestimation of the precision of INR estimates. This may explain some of the discrepant and “inexplicable” INR values, which are frequently reported in the scientific literature.
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Abbreviations

List of abbreviations with relevant page numbers in parentheses.

ANOV A Analysis Of Variance
BLUE Best Linear Unbiased Estimator (47)
CV Coefficient of Variation
DEKS Danish Institute for External Quality Assurance in Hospital Laboratories; Dansk Institut for Ekstern Kvalitetssikring for Sygehuslaboratorier (36)
ICC International Calibration Constant (20)
ICR International Calibrated Ratio (20)
INR International Normalized Ratio (22, 36, 68)
IRP International Reference Preparation (18)
ISI International Sensitivity Index (22, 36, 68)
MNPT Mean Normal Prothrombin Time (35)
OAT Oral Anticoagulant Therapy (1)
OLS Ordinary Least Squares (23)
PP Prothrombin Percentage (7)
PT Prothrombin Time (5, 35)
PT ratio Prothrombin time ratio (36, 76)
PT system Prothrombin time system (35)
PIVKA Proteins Induced by Vitamin K Absence or antagonists (14)
se  Standard error (of the mean) (46)

SOP  Standard Operating Procedure (36, 44)

TSR  Thromboplastin Sensitivity Ratio (10)

WHO  World Health Organization