Estimation of Preanalytical Uncertainty in Clinical Chemistry

Marit Sverresdotter Sylte

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

2013

Scientific environment

This work was carried out at Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway and Department of Clinical Science, Faculty of Medicine and Dentistry, University of Bergen, Norway under supervision of Professor Bjørn J. Bolann, in close collaboration with co-supervisor Bio-statistician Tore Wentzel-Larsen at the Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway, Norwegian Centre for Violenced and Traumatic Stress Studies, Oslo, Norway, Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Oslo, Norway.

The quality of your life is in direct proportion to the amount of uncertainty you can comfortably live with.

Tony Robbins (1960 -)

To my father, Kristin, Aksel and Ingvild

Contents

Acknowledgements

This work has received financial support from: The Western Norway Regional Health Authority ("Samarbeidsorganet"), Laboratory of Clinical Biochemistry, Haukeland University Hospital, Centre for Clinical Research, Haukeland University Hospital, and The Norwegian Society of Engineers and Technologists (NITO).

I am very grateful to and would like to thank:
Professor Bjørn J Bolann, principal supervisor, for introducing me into the academic way of thinking, giving me great freedom, while still always being available, for being a good discussion partner, for follow up of the work and being "the decision-making authority", and for your good sense of humour. I greatly appreciate our work together.

Bio-statistician and supervisor Tore Wentzel-Larsen for excellent cooperation with statistical modelling, and answering many questions, always having the assumptions ready, reading through many drafts, with Argus eyes, and giving the quickest and best feedback possible. Your ability to acquaint and cope with our laboratory practice, has been valuable and decisive for the statistical modelling.

Head of the department Anne Grete S. Thue, and previous Head of the department Sverre Sandberg, for giving me financial support and great confidence. I am proud of the department, and the important role research has in the routine practice. I am very glad for having had the opportunity to carry out the work that goes into this thesis.

The outpatient clinic for providing facilities for collecting the blood samples, the medical technicians at the Section for Automatic analysis for always with a smile, running all the samples, and my colleagues for general help and support.

Regina Kufner Lein at the Medical Library at the University of Bergen Library for helping me with reference lists and literature search.

My closest colleagues, Astrid-Mette Husøy and Marit Økland, for being patient and spreading happiness every day at work.

Last, but not least, my family and friends; especially my dear father, and my dear children Kristin, Aksel and Ingvild.

Bergen, March 2013

Marit Sverresdotter Sylte

Abbrevations

ALP	alkaline phosphatase
ALT	alanine aminotransferase
BIPM	Bureau International des Poids et Mesures
CI	confidence interval
CK	creatine kinase
CLIA	Clinical laboratory improvement amendments
CLSI	Clinical and Laboratory Standards Institute
EQA	external quality assessment
GGT	γ-glutamyltransferase
GUM	Guide to Expression of Uncertainty in Measurement
H-index	hemoglobin index
ISE	ion-selective electrode
ISO	International Standards Organization
IVD	in vitro diagnostic
LD	lactate dehydrogenase
LKB	Laboratory of Clinical Biochemistry
LME	linear mixed effects models

MSE	mean square error
RCV	reference change values
RMSE	root mean square error
RST	rapid serum tube
SST	serum separation tube
TC	total cholesterol
TG	triglycerides

Abstract

Preanalytical uncertainty is attributable to variations in blood sample collection and sample handling before analysis. The aim of this study was to establish a modelling framework for estimating preanalytical uncertainty. There is a need for standardization on which uncertainty sources that should be included, and how the preanalytical uncertainty should be estimated. In Paper I, an uncertainty budget was established based on differences in paired data between a standard method for handling blood samples and alternative methods used in current practice, considering the distribution of alternative methods. In Paper II and III, linear mixed-effects models were used to estimate the between-venipuncture SD, the preanalytical SD (excluding the between-venipuncture SD), and the measurement repeatability when the phlebotomy and the sample handling were performed optimally, and any difference in preanalytical SD and fixed effects, between transporting blood samples in a pneumatic tube system vs manual delivery, using different needles or tubes, and mixing methods. When the combined biases from the uncertainty budget in Paper I and the significant biases between different treatments in Paper II and III were compared with defined quality specifications for analytical bias, glucose was the only analyte falling outside the quality specifications. Prolonged clotting and storage time were the greatest contributors to the bias for glucose, and the significant mean difference between SST vs RST tubes shows that choice of tube is important. The preanalytical SDs (excluding the between-venipuncture SD) for LD and potassium for optimally treated samples, were significantly higher than the measurement repeatability SDs, but for glucose, the between-venipuncture SD was the dominant source of variation. For most analytes, the preanalytical SDs were about the same in both studies II and III, indicating that the preanalytical variations are little influenced by different preanalytical handling. We have developed two models that can be used to estimate preanalytical uncertainty in clinical chemistry laboratories. Estimation of preanalytical uncertainty may improve diagnostic quality and patient treatment.

List of publications

This thesis is based on the following original papers, referred to in the text by their Roman numerals.

Paper I

A model for an uncertainty budget for preanalytical variables in clinical chemistry analyses.
*Rynning M, Wentzel-Larsen T, Bolann BJ. Clinical Chemistry (2007) 53:1343-1348.

*Change of name to Sylte.

Paper II

Estimation of the minimal preanalytical uncertainty for 15 clinical chemistry analytes.
Sylte, MS, Wentzel-Larsen, T, Bolann BJ. Clinical Chemistry (2010) 56:1329-1335.

Paper III

Random variation and systematic error caused by various preanalytical variables, estimated by linear mixed-effects models.
Sylte, MS, Wentzel-Larsen, T, Bolann BJ. Clinica Chimica Acta (2013) 415:196-201.

1. INTRODUCTION

The uncertainty of the result of a measurement reflects the lack of exact knowledge of the value of the measurand (1). All test results produced in clinical chemistry laboratories are encumbered by uncertainty, and are estimates of true values. Knowledge of the result variability is required if results are to be meaningfully compared with previous results from the same patient, or with clinical decision limits (2).

A laboratory mistake can be any defect during the entire testing process, from ordering the tests to reporting the test results (3). A study on the frequency and types of mistakes in a laboratory found that 68% was caused by preanalytical mistakes, 13% by analytical mistakes, and 19% by postanalytical mistakes (3). In the preanalytical phase samples could be exposed to mistakes such as misidentification of patients and specimens, blood collections could be performed by personnel less skilled, specimen material could be wrong, samples could be missing, and quantity and quality of the specimen could be inappropriate (4). Misidentification of the patient is a serious mistake that may result in wrong medical treatment for the patients involved.

If a series of blood samples is collected over time from one individual for a particular laboratory test, the results will vary randomly (5). This random variation consists of intra-individual biological variation, preanalytical variation and analytical variation. The intra-individual biological variation, also called within-subject variation, causes a random fluctuation around a homeostatic setting point due to natural biological factors (5). In addition, physiological factors such as age, activity before sampling, food ingestion, menstrual cycle, the patients' posture during the phlebotomy, pregnancy, smoking, and the time of the day for the collection may also influence test results (6).

Measurement uncertainty should not be confused with production error or mistake (7). In laboratory medicine, uncertainty sources are commonly grouped as affecting the premeasurement, measurement and postmeasurement phases (2). Basic knowledge of

implementation, verification, and maintenance of laboratory equipment is essential for producing accurate and precise test results (8). Possible changes in test results should reflect the changes within the patient, and not represent changes in the trueness or in the imprecision of the method (5). A possible bias should not move the patient from one diagnosis group to another (9). Automation, standardization and technical development have significantly improved the analytical quality and reliability of laboratory results (4). The trueness and analytical precision of analyzing methods are known and verifiable for most analytes. Technological advances such as barcode technology, primary tube processing, serum indices, delta check technology, and volume/clotting/bubbles sensors, have increased the quality of test results (10).

The phlebotomy itself, type of device used to collect the blood, handling of the samples after the collection, transport and storage of the specimens may influence the measured concentration of components in blood (11). Preanalytical handling demands detailed knowledge about recommended tourniquet time, choice of tube, filling and mixing, clotting time, centrifugation speed and time, transport, and storage. Since premeasurement uncertainties have been difficult to estimate, common practice has been to minimize, where possible, the uncertainties by implementing standardized procedures for specimen collection, and handling of the blood samples (2).

There are several evidence-based guidelines in laboratory medicine covering topics within both preanalytical, analytical, and postanalytical elements (12). Guidelines should be based on the critical evaluation and systematic review of literature, and explicitly state the strenght of evidence supporting each recommendation (12). An international organization provides consensus guidelines and standards for patient testing and health care services, and several standards give descriptive and stepwise procedures for improving preanalytical handling of blood samples (6;13-15).

In laboratory medicine, even if there is substantial emphasis on how different preanalytical and analytical factors influence test results, the total uncertainty of test results, where both the preanalytical and analytical uncertainty are combined, for most

biochemical components is unknown. In a collective paper with synopsis of lectures from a conference in preanalytical phase, different preanalytical errors and strategies to increase the quality are discussed (16), but estimating the preanalytical uncertainty is not emphasized.

Sampling is part of nearly all chemical measurement. It has been argued that the uncertainty caused by sampling has been ignored, and that sampling protocols are not validated compared with analytical methods (17). Knowledge of both sampling and analytical uncertainty is necessary, in order to make the correct decisions on the test results, and to ensure that resources are distributed optimally (17). The fundamental sampling error is stated to be the minimum error of an ideal sampling procedure (18). It is proposed that the total sampling error can be divided into errors of incorrect sampling and errors of correct sampling. Incorrect sampling error causes biases and increases the total variance in a unpredictable way, and trying to estimate it would give estimates that cannot be generalized. Errors of correct sampling may be quantified as measurement uncertainty arising from four sources of errors, the sampling and analytical precision, and the sampling and analytical bias (18).

According to ISO standard 15189 for accreditation, the uncertainty of results should be estimated when it is relevant and possible (19). Estimating the preanalytical uncertainty is a challenge because the blood samples are exposed to many different treatments in current practice, although there are recommended procedures for optimal treatment. As a result of this complexity, estimation of preanalytical uncertainty demands a model that can estimate both random variations and systematic deviations caused by the different preanalytical treatments. Identification of the factors that contribute to the variability of repeated measurements results, may provide valuable insights into the reliability of results, and potential means for improvement (2).

In several preanalytical studies, the effect of preanalytical variables was estimated as systematic deviations between a specific preanalytical handling practice, and different alternative practices (20-22). Some researchers have estimated the preanalytical

variation for a specific preanalytical practice, without distinguishing between random and systematic effects (23-25). The combined uncertainty for glucose, including the uncertainty from patient preparation, position of the patient, sampling and the measurement procedure, was estimated based on assumptions of plausible intervals and rectangular distributions (26). In another study, paired data were used to calculate the uncertainty from specimen collection, effect of delay in pretreatment phase and transportation (27). The preanalytical uncertainty components from the different experiments, were combined with data on analytical variation, and biological variation (27). Some researchers combined the standard uncertainties from the measurement, preanalytical variation and intra-biological variation (28). In this study, the uncertainty components were quantified using observations of the measuring system, and information from calibration certificates, instrument specifications and literature (28).

To assess the reliability of a test result, and to compare test results with each other, a generally accepted procedure for evaluating and expressing the uncertainty is desirable (1). There is a need for standardization on which uncertainty sources that should be included in the preanalytical uncertainty, how the preanalytical uncertainty should be estimated, and what the assumptions are.

This thesis introduces two models that can be used by the laboratories to estimate preanalytical uncertainty for biochemical components. In the following, the introduction includes some theory of uncertainty, and a description of the preanalytical variables involved in our studies.

1.1 Random and systematic error

Traditionally, we differentiate between random and systematic errors. Error is defined as “measured quantity value minus a reference quantity value” (7).

Random measurement error is defined as the “component of measurement error that in replicate measurements varies in an unpredictable manner” (7). The random errors are

caused by many variables both in the preanalytical and the measurement process. In the preanalytical process they may be caused by variation in the phlebotomy, clotting time, centrifugal force, storage conditions, transport of the samples, etc.

Systematic measurement error is defined as a "component of measurement error that in replicate measurements remains constant or varies in a predictable manner" (7). The systematic measurement error, and its causes, can be known or unknown. Systematic deviations, also called biases, influence the test results, and push the test result in one definite direction. Over a period of time, some short-term biases may be regarded as random variation (2). A correction can be applied to compensate for a known systematic measurement error. In preanalytical treatment, a systematic error can be introduced as a result of treating the blood samples differently than recommended, such as expanding the clotting time, using other tubes than the standardized one, and choosing another way of transport.

Measurement repeatability (the within-run precision) is defined as "measurement precision under a set of repeatability conditions of measurement" (7). The repeatability precision is defined as "condition of measurement that includes the same measurement procedure, operators, measuring system, operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time" (7). Measurement reproducibility (the between-day precision) is defined as "measurement precision under reproducibility conditions of measurement" (7). The reproducibility precision is "condition of measurement, that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects" (7).

1.2 Guide to Expression of Uncertainty in Measurement

Uncertainty of measurement is defined as "a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used" (7). The Guide to Expression of Uncertainty in Measurement

(GUM) provides guidelines for expressing uncertainty in measurement, including those for converting the uncertainty estimates into standard form, combining them, and calculating the combined uncertainty (1). The aim of GUM is to harmonize the different practices for estimating and reporting uncertainty of measurement (29). Measurement uncertainty may include components arising from systematic effects, and components evaluated by Type A and Type B evaluations of measurement uncertainty (1). Both types of evaluation are based on probability distributions, and the uncertainty components are quantified by variances (1). Type A evaluation of uncertainty is based on statistical analysis of series of observations (1). The type A experimental standard deviation characterizes the variability of the observed values or their dispersion about their mean (1). Type B standard uncertainty is evaluated by scientific judgement based on available information on possible variability, such as data provided in certificates, handbooks, manufacturer's specifications (1). When the number of observations is limited, a type B evaluation may be as good as a type A evaluation. The distribution of the errors in a Type B uncertainty is claimed to be uniform, because sometimes, all that is known are the end-points of the interval wherein the quantity varies (30).

Many guidelines or standards are based on the GUM guidelines (7;18;31;32). A bibliography on uncertainty presents several general and specific applications documents derived from the GUM over many years (33). According to GUM, the evaluation of uncertainty is neither a routine task nor purely mathematical, it depends on detailed knowledge of the nature of the measurand and of the measurement (1). Therefore, construction of an uncertainty model can be complicated (34).

An application of a model, based on GUM, for evaluating the uncertainty of a measurement result has been demonstrated, in order to harmonize the uncertainty evaluation process (35). Data from method validation studies, internal quality control and external quality assessment schemes (EQA) are used in the evaluation (35). A critique of GUM is that too narrow uncertainty intervals will be presented to the clinicians and consequently be misleading, because unexplained outliers that are not

unusual in diagnostic assays, due to interfering substances, will not be included in the intervals (36).

1.3 Quality specifications based on biological variation

Laboratory test results are used for many purposes such as diagnosis, case finding, screening, and monitoring (5). Quality specifications should ensure that these clinical purposes can be achieved (5). Quality specifications for total allowable analytical error can be derived from evaluation of the effect of analytical performance on clinical outcome, biological variation, clinician's opinions, official regulatory bodies, external quality assessment and state of the art (37). It is well documented that objective quality specifications are necessary in method evaluation and quality control (5). Quality specifications for analytical precision and trueness derived from biological variation have been set (38). An updated list of analytical quality specifications is available on Westgard's homepage (39).

According to these specifications, for monitoring a patient's condition, the analytical variation (CV_A) has to be below half of the within-subject biological variation (CV_I) (39). For screening, and diagnosis, when a cut-off point is used, and as related to reference intervals, the analytical bias (B_A) should be below a quarter of the square root of the squared within-subject (CV_I) plus the between-subject biological variation (CV_G) (39):

$$CV_A \leq 0.5 \times CV_I$$

$$B_A < 0.25 \times (CV_I^2 + CV_G^2)^{1/2}$$

1.4 Treatment of biases

Systematic and random error cannot be eliminated, but it can often be reduced. GUM states that if the systematic effect is significant in size relative to the accuracy of the

measurement, a correction can be applied to compensate for the effect, and that "it is assumed that the result of a measurement has been corrected for all recognized significant systematic effects, and that every effort has been made to identify such effects" (1). Although it is recommended to correct for biases, this may not always be practical (40). It is proposed, that when there are several sources of uncorrected biases, the biases should be added, and the combined bias should be stated together with the combined standard uncertainty (40). If these biases are not independent, the degree of overlap of the biases should be estimated and subtracted, to avoid doubly counting biases (40). According to GUM, measurements results should be corrected for the bias, and the uncertainty in the bias correction should be included in the combined standard uncertainty. In the proposed approach a complete uncertainty statement should include the combined standard uncertainty (biases are corrected), a statement of the bias value, and an expanded uncertainty including the effect of the bias (40).

1.5 Traceability, trueness, accuracy

Traceability, trueness and accuracy are related terms used in analytical measurements. These terms may also be used in the preanalytical field. A metrological traceability chain is defined as "sequence of measurement standards and calibrations that is used to relate a measurement result to a reference" (7). The trueness is defined as "closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value" (7), while accuracy is defined as "closeness of agreement between a measured quantity value and a true quantity value of a measurand" (7).

Establishing metrological traceability satisfies the basic requirements of evidence-based laboratory medicine (41). The manufacturers have to document the metrological traceability for commercially available methods (41). The methods should preferably be traceable to a primary reference material or a reference method, and the goal is to assure that test results are accurate and comparable over time (41). Specifically,

because of a standardization program for glycohemoglobin, harmonization of test results between methods and improved analytical precision have become possible (42).

Shifts in bias or changes in trueness will influence on the number of patients classified having a disease according to established decision thresholds, and thereby affect both medical outcome and healthcare costs (9). Allowing a bias on e.g. ± 1% for cholesterol, will cause approximately ± 6% change in the number of patients passing decisions thresholds (9). There are analytical performance standards for individual analytes such as the CLIA criteria (43), which involve quality standards for all laboratory testing to ensure the accuracy, reliability and timeliness of patient test results, regardless of where the test is performed. In EQA, the accuracy of a laboratory's analysing method is determined by comparing the test result for each analyte with a target value defined by e.g. the mean of the participants' responses, or value established by reference method. The acceptance limits for quantitative tests in EQA are related to the analytical variance. The limits are set to ensure that the test results are clinically useful, and that the performance compared with other laboratories is acceptable. A high between-laboratory variability may cause wider acceptance limits in EQA than what is clinically desirable (44). In order to allow 95% of laboratories to meet an EQA challenge, a consensus on minimum level of analytical performance has been set by Spanish scientific societies organizing EQA (45).

1.6 Statistical methods

There are two approaches to the estimation of uncertainty, the "bottom-up" and the "top-down" (18). The "bottom-up" approach quantifies the sources of uncertainty individually, and then uses a model to combine them. In contrast, the empirical or "top-down" approach involves some level of replication of the whole measurement procedure to give a direct estimate of the uncertainty (18). The "top-down" approach can be used to estimate the uncertainty from one or more effects, or classes of effects

(18). The bottom-up procedure may be useful during method development, and the top-down approach for method verification (2).

By definition, an uncertainty budget is a "statement of a measurement uncertainty, of the components of that measurement uncertainty, and of their calculation and combination" (7). In study I (Paper I), an uncertainty budget was modelled by summing up the expected individual biases and variances in paired data, between current practice and the standard method for each uncertainty source, considering the distribution of the alternative methods. The modelling in study I is a "bottom-up" approach. Doing modelling is about making proper simplifications and assumptions, and the assumptions must be validated (34). Study II and III (Paper II and III) are "top-down" approaches, where linear mixed-effects models were used to identify different sources of variation. Mixed-effects models allow the use of clustered multilevel data, and separate estimates of fixed and random effects (46). By collecting blood into several tubes from both arms from several individuals, and analysing the specimens in duplicate, the data set reflected four sources of variability: The between-subject, the between-venipuncture, the between-tube defined as the preanalytical variation, and the measurement repeatability. Duplicate analyses of control samples are a recommended method for estimating analytical precision (47).

Another choice of statistical method is Empirical Bayesian models that could be used as an alternative statistical method to linear mixed-effects models (46). On the contrary, analysis of variance is not adapted to data with cluster structure, and multivariate analysis of variance is built on analysis where several analytes are dependent variables in the same analysis, and consequently based on additional assumptions that not necessarily need to be fulfilled. As to repeated measures analysis of variance, it is an older method for clustered data, having weaknesses that linear mixed-effects models do not have. Linear mixed-effects models do e.g. not demand balanced data, and build on fewer assumptions (48).

1.7 Preanalytical variables

1.7.1 Sample collection

Tourniquets should not be used at higher pressures than 40 mmHg (6), and not longer than one minute (11). Long tourniquet application time may result in increase of the concentration of macromolecules, blood cells and compounds bound to proteins, while the concentration of low molecular analytes are less influenced, because fluid and low molecular compounds are moving from the vein into the interstitium (11). Repeated clenching and unclenching of the fist should be avoided.

Needles and winged blood collection sets are color-coded according to the size of the inside diameter of the needle (6). The 23-gauge baby-blue butterfly needle indicates a smaller diameter than the 21-gauge straight green needle. Collecting blood by using the green straight needles is considered gold standard in this work, and is the most sold needle (personal communication) according to the Norwegian manufacturer PULS. The long flexible tube of the butterfly needle has a dead volume of about 0.5 mL. Of the needles used at the pediatric clinic at Haukeland University hospital 50-60% are blue butterfly needles. At the outpatient clinic at the main hospital laboratory about 20-30% is butterfly needles, equally distributed on blue and green butterfly needles. A study found no bias when comparing butterfly needles having different gauge for analytes such as calcium, CK, LD and potassium (20). When straight needles were compared with butterfly needles (both 21-gauge needles), there were found no significant differences in the results for several analytes including ALP, calcium, CK, LD, and potassium (49).

By venipuncturing both arms in study II and III, the between-venipuncture variation could be estimated, in addition to the fixed effects representing any systematic differences between arms, and between types of needles. The tubes should be completely filled, and immediately mixed gently by 5-6 inversions to disperse the clot activator. Vigorous mixing should be avoided. A study on Li-heparin gel tubes showed

a small significant increase of LD and H-index in samples subjected to instant mixing directly after the phlebotomy compared to samples without mixing (21).

1.7.2 Sample handling

To ensure test reliability, well-trained personnel using appropriate devices is needed when collecting blood (50). Gel tubes are widely used to separate serum from clotted whole blood. The gel has a controlled viscosity and a specific gravity which make the gel move between the serum and the clot during the centrifugation. To avoid interference from the gel it is important to follow the manufacturer's recommendation for temperature while storing the tubes, and using the recommended centrifugal force. Blood collection tubes should be validated in order to give accurate and precise test results (15). The draw should be within 10% of the stated draw, and the amount of additive should be within the range specified by the manufacturer (14).

The pre-centrifugation phase is the time interval between specimen collection, and centrifugation (13), called the clotting time in our studies. Complete clotting normally occurs within 30 to 60 minutes at room temperature (13). Gel tubes with thrombin such as the rapid serum tubes (RSTs) only need 5 min to form a fibrin clot, while gel tubes with silica clot activator such as the serum separation tubes (SSTs), need 30 min. For plain tubes of glass without additives, 45-60 min clotting time is recommended. In clinical research studies, it is recommended to use the same kind of tube from the same manufacturer, in order to reduce the uncertainty (50).

Modern centrifuges have swing-out rotors, are temperature controlled, and the speed of the centrifuge can easily be adjusted either by g-force or speed of rotation. Operational verification activities of the centrifuges such as speed control and timer are strongly recommended (8). It is recommended to follow the specifications of the manufacturers of the tubes regarding centrifugation speed and time.

To prevent disturbance during the clot formation, the tubes should be kept in a vertical, closure-up position (13). Blood attached to the tube closure may result in hemolysis in

the serum. Serum should be separated from the erythrocytes as soon as possible (13), but studies have shown that many analytes were unaffected by cells contact time for 24 to 72 hours at room temperature (13;51;52). However, the concentration of glucose decreased, and LD and potassium increased after 24 hours contact with the clot (52). Magnesium is shown to increase 11% after 24 hour delay of centrifugation (53). For glucose, LD and potassium it is stated up to two hours stability for uncentrifuged specimens (13). Glycolysis is the conversion of glucose or other hexoses into lactate or pyruvate (54). Glycolysis decreases serum glucose by approximately 5 to 7% in 1 hour in normal uncentrifuged coagulated blood at room temperature (54).

Temperature and time of storage may influence the stability of biochemical components in serum. The serum should be stored at 2–8°C when not analyzed within 8 hours after the phlebotomy (13). Serum can be stored on gel for up to 48 hours at 4 ^{o}C for most analytes with the exception of some drugs (13). In non-hemolyzed serum, separated from the erythrocytes, the glucose concentration is stable for 8 hours at room temperature (54). Manufacturers should be requested for stability data since different analysing methods may have different stability requirements for the same measurand (13).

1.7.3 Pneumatic tube transport

Benefits of pneumatic tube systems are improved efficiency and productivity. Several researchers have studied the effects of transporting blood samples by pneumatic tube systems (22;55-60). Shaking and gravity forces may influence the specimens, e.g. by hemolysis, and thereby change the concentration of some analytes (22;55-59). A positive relationship is shown between the speed of the pneumatic tube system, and increased changes in concentration of potassium and LD (22). Transporting unclotted blood and clotted blood have been shown to give the same effect on LD, but transporting incompletely filled tubes increases the effect on LD further (59).

1.7.4 Hemolysis

Hemolysis is release of hemoglobin and other intracellular components from erythrocytes into plasma/serum (11). Hemolysis may be classified as *in vivo* and *in vitro* hemolysis. *In vivo* hemolysis may result from intravascular erythrocyte destruction (13), while *in vitro* hemolysis from the phlebotomy and sample handling before analysis. Fragile veins, collection of blood from a hematoma site, prolonged tourniquet time, equipment that may lead to turbulent blood path such as butterfly needles, and vigorous mixing are examples on blood collection that may result in *in vitro* hemolysis (61).

The effects of hemolysis are 1) increased intracellular constituents such as LD in serum, 2) optical interference, and 3) interference with the reaction mechanism of the assay (11). Interference may occur even by low concentration of hemoglobin (11). The concentration of potassium in erythrocytes vs normal plasma is 23:1, and for LD 160:1, respectively (62). Hemolysis is shown to be the leading cause of unsuitable specimens in clinical laboratories (61), and may be a suitable indicator for preanalytical quality (63).

1.7.5 Literature search

The general literature search was completed at the end of February 2013. Search criterias in Pubmed (My NCBI what's new results from the National Center for Biotechnology Information at the U.S. National Library of Medicine) were “clinical chemistry uncertainty”, “preanalytical phase”, “preanalytical”, “blood tubes phlebotomy”, “uncertainty budget”, “stability clinical chemistry analytes storage”, “clinical chemistry quality specifications”, and “metrological traceability”.

A specified literature search for using linear mixed-effects models in estimating uncertainty in clinical chemistry was performed 18.12.2012. Only one study (64), from the search in Web of Science, had some similarity to our studies II and III, concerning the approach, design, method, subject area and research in question.

Detailed description of the literature search 18.12.2012:

Search in Pubmed: Search queries as **"**Laboratories, Hospital" [Mesh] OR (Blood Specimen Collection OR Phlebotomy OR preanalytical) gave 21298 matches, while the search query (linear mixed effect model) gave 1527 matches. The combination gave 7 matches. Paper II was not among the 7. None of matches showed similarity to study II and III.

Search in EMBASE: The search query ("linear mixed effect* model*") gave 1618 matches. Search queries (analytical error) OR (blood analysis) OR (blood sampling) OR (preanalytical) OR (laboratories) OR (laboratory) gave 349881 matches. The combination gave 132 matches, where one of them was Paper II. Evaluated from the titles and the abstracts, the other 131 matches were not of current interest.

Search in Web of Science: The search query ("linear mixed effect*model*") gave 1468 matches. Search queries as (laborator*) OR (preanalytical) OR ("blood specimen") OR (phlebotomy) OR ("blood analys*") OR ("blood sampling") OR ("analytical error") gave 503155 matches. The combination gave 41 matches, where one of them was Paper II. Based on evaluating the titles and the abstracts, only one of the papers may be of current interest (64), as mentioned above. The researchers examined the effect of time, tube, anticoagulant type, on serum and plasma profiles within low-molecular-weight proteome (64). Blood was collected into several tubes from 3-6 individuals in each project, and the tubes were processed differently further on. Cluster analysis was used to identify samples with similar peak profiles. Mean intensities of the different groups were estimated and compared to a basis point. Compared with our studies II and III, neither random effects nor confidence intervals were estimated.

Search in MathSciNet: The search query "linear mixed effects model*" gave102 hits. The papers were within mathematical statistics.

Search in Jstor Statistics: The search query ((“linear mixed effects model*”)) AND (laborator* OR preanalytical OR blood) gave 150 hits. Among these 150 matches we were not able to find any appropriate paper with preanalytical or laboratory focus.

2. AIMS

The overall aim of this thesis was to establish a modelling framework for estimating preanalytical uncertainty in clinical chemistry analyses. The specific aims were:

Paper I

For each uncertainty source, summing up the expected biases and variances in paired data between a standard method for handling blood samples and alternative methods used in current practice, and considering the distribution of alternative methods, develop an uncertainty budget for preanalytical variables in clinical chemistry analyses. The uncertainty budgets should include the uncertainty added to test results by using different kinds of blood tubes and instruments, prolonged clotting time and centrifugal force, and delays in measurement.

Paper II

By linear mixed effects-models, detect any difference in preanalytical variation when blood is collected into two different types of tubes, and estimate the between-venipuncture variation, and the preanalytical variation (excluding the between-venipuncture variation), and the measurement repeatability, together with any fixed effects, when the phlebotomy and the sample handling are performed optimally for 15 clinical chemistry analytes.

Paper III

By linear mixed-effects models determine whether specific, preanalytical treatments increase preanalytical variation and bias test results compared with optimal treatment for 21 clinical chemistry analytes. The following treatments are examined: Transporting blood samples in a pneumatic tube system vs manual delivery, collecting blood using 23-gauge butterfly needles vs 21-gauge straight green needles, and mixing blood samples by only one inversion vs five inversions.

3. MATERIALS AND METHODS

Table 1 presents the preanalytical variables examined, the participants included, the data structure, statistics and outcome in Paper I, II and III.

Table 1. Preanalytical variables examined, participants included, data structure, statistics and outcome in Paper I, II and III.

Paper	Preanalytical variables	Participants	Data structure	Statistics	Outcome
I	Blood tubes	Hospitalized patients n=33 + 34	Paired observations[a] Discrete distribution	Own model[b]	Bias (SD[c])
	Clotting time	Hospitalized patients n=45	Paired observations Continuous distribution	Own model	Bias (SD)
	Centrifugal force	Hospitalized patients n=28	Paired observations Continuous distribution	Own model	Bias (SD)
	Storage time	Hospitalized patients n=31	Paired observations Continuous distribution	Own model	Bias (SD)
	Instruments	Serum from the routine n=500	Paired observations Discrete distribution	Own model	Bias (SD)
II	Optimal treatment	Volunteers n=20	Hierarchical cluster data	Linear mixed-effects models	Bias (95% CI) SD (95% CI)
III	Ways of transportation (experiment 1)	Outpatient clinic n=30	Hierarchical cluster data	Linear mixed-effects models	Bias (95% CI) SD (95% CI)
	Needles and mixing methods (experiment 2)	Outpatient clinic n=30	Hierarchical cluster data	Linear mixed-effects models	Bias (95% CI) SD (95% CI)

[a] Paired observations between the alternative method and the standard method.
[b] An uncertainty budget is modelled by combining expected biases and variances between standard method and current practice from each uncertainty source.
[c] SD of the differences between the paired data.

3.1 Ethical considerations

All studies were performed in accordance with the Helsinki declaration. Study II and III were approved by the Regional Committee for Medical and Health Research Ethics, Western Norway (REC no 022.28). Study I was a quality assurance project, and quality assurance and evaluations which is part of the health services, and technical and methodological scientific development using anonymously biological materials, is exempted from approval by Regional Committee for Medical and Health Research Ethics, Western Norway. Written informed consent was obtained from all participants in study II and III, while in study I informed oral consent was obtained. In study II and III, the test results were assessed by the medical doctor executive of the project. In all studies the name, and date of birth was removed, and the samples were analysed anonymously. The sample materials were discarded after measuring. The following documents are enclosed in the appendices: The approval documents from the Regional Committee for Medical and Health Research Ethics (no. 1), the Norwegian Social Science Data Services (no. 2), the Norwegian Directorate of Health (no. 3), and the enquiry and consent declaration for participation (no. 4).

3.2 Participants and sample collection

In study I, the single blood tube needed from each patient was collected from consenting hospitalized patients in the course of routine blood collections for tests already ordered by the patients' physicians. The patients were conscious, and able to understand the enquiry. Most of the phlebotomies were done by the same medical technician. Paper 1 consisted of 5 subprojects. Four of five subprojects involved phlebotomy on approximately 30 patients in each, while one involved 500 serum samples previously analysed on two different Roche Modular Analytics SWA instruments by photometric methods (Roche Diagnostics GmbH) in our department.

In study II, blood was collected from 20 non-fasting healthy volunteers employed at our laboratory. There were no specified inclusion or exclusion criteria. The mean venipuncture duration for both arms was 3 min (range 2 – 7 min).

In study III, blood was collected from 30 non-fasting, consecutively recruited patients at our outpatient clinic, separately for experiment 1 and experiment 2. The sampling was stratified as approximately equal number of out-patient men and women above 18 years of age. The mean venipuncture duration for both arms was 3.5 min (range 2.5 – 6 min). To limit the amount of blood collected, patients with only a few routinely ordered tests were chosen. In study II and III the same medical technician performed all phlebotomies with the participant remaining in a sitting position for approximately 10 min at ambient temperature between 9 AM – 1 PM. The tourniquet was loosely fastened and released after < 1 minute, as soon as blood appeared. Repeated clenching and unclenching of the fist was not allowed. Table 2 shows the different blood tubes used in the studies.

Table 2. Vacutainer blood tubes (Becton Dickinson, USA) used in the studies.

Paper	Type of blood tube	Tube material	Additives
I	Serum separating tube (SST)	Glass	Gel + silica clot activator
	SST II Plus	Plastic	Gel + silica clot activator
	Plain	Glass	None
II	Rapid serum tube (RST)	Plastic	Gel + thrombin-based medical clotting agent
	SST II Advance	Plastic	Gel + silica clot activator
III	SST II Advance	Plastic	Gel+ silica clot activator

Power analysis can be used to calculate the sample sizes necessary to detect a specified difference when the error variance is known (or can be guessed at) (65). We did not perform power calculations in any of our studies. In study I we did not have a priori data on the SD of the differences in the paired data between the standard method and

the alternative method, while in study II it was not possible to do power calculations, because we did not have a priori SDs for the between-individual SDs, the between-venipuncture SDs, the preanalytical SDs (excluding the between-venipuncture SD) or the measurement repeatability SDs.

In study III, the sample size was increased from n=20 to n=30, because of a more heterogeneous population than in study II. We did not have a priori SDs for the new treatments and levels in study III, such as the between-venipuncture SDs when using both butterfly and green needles, preanalytical SDs (excluding the between-venipuncture SD) for suboptimally treated samples, or measurement repeatability SDs for duplicates from test results from patients.

3.3 Analytical methods

Laboratory of Clinical Biochemistry (LKB) produced about 6.2 million test results in 2012. The analytes examined in study I, II and III were among the most generally ordered at our laboratory (Figure 1). In study I four analytes were examined, in study II 15 analytes, and in study III 21 analytes (specified in Table 4). The analyses should not be too expensive, since the studies involved many specimens and duplicates analysis. The clinical chemistry measurements were performed at Roche Modular Analytics SWA (serum work area) on P800 module instruments by photometric methods from Roche Diagnostics GmbH. The photometric methods are detailed in the papers. The electrolytes were measured with ISE (ion-selective electrode) indirect method, while folate was measured with the Elecsys competitive method Folate III on E170 modules (Roche Diagnostics GmbH). The hemolysis was measured by a photometric method as the hemoglobin index (H-index), where 100 H-index units correspond to a hemoglobin concentration of about 0.06 mmol/L (0.1 g/dL). The analysing methods have good analytical precision and trueness. The specimens were analysed anonymously and randomly in duplicates within the same analytical run to ensure the same measuring conditions.

At LKB, the standard procedure is to reject test results for LD and folate when the H-index is above 50, and for potassium when the H-index is above 65.

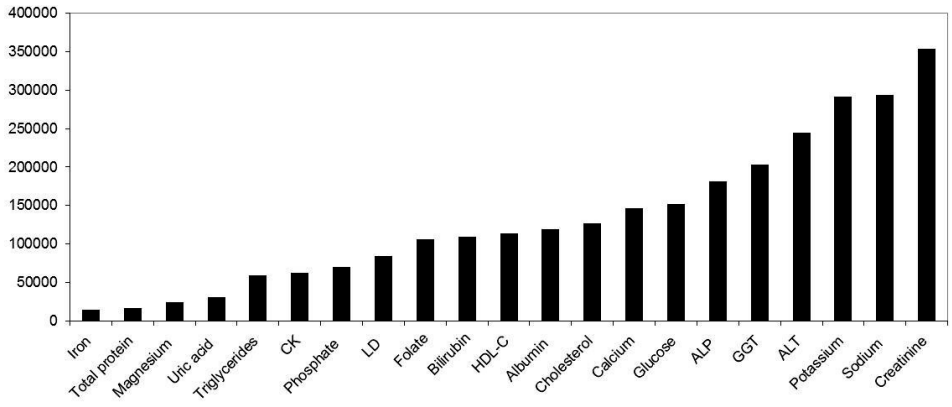

Figure 1. The figure shows the analytes examined, and the number of analyses performed at LKB in 2012.

3.4 Paper I

3.4.1 Specimen handling

For each uncertainty source, we specified a standard (recommended) method of preanalytical treatment and alternative methods of treatment, which were within clinical practice (Figure 2). Each of the 5 uncertainty sources (*a–e*) was studied separately by paired observations between the alternative methods and the standard method. The alternative methods are used to a variable extent in current practice.

a) Different blood tubes: One SST tube (standard method) and one plain tube (alternative method) were collected from each patient (N=33), and one SST

(standard method) and one SST II Plus tube (alternative method) were collected from each patient (N=34). All pairs of tubes had equal clotting time (between 45 and 120 min) and were centrifuged at 1300*g*, except the SST II Plus tubes that were centrifuged at 1850*g* as recommended.

b) Clotting time: Two SSTs were collected from each patient (N = 45). One of the paired SSTs was centrifuged as recommended after 40–70 min clotting time (standard method), and the other after 120–150 min (alternative method).

c) Centrifugal force: Two SSTs were collected from each patient (N = 28). One of the paired SSTs was centrifuged at 1300*g* (standard method) and the other at 2350*g* (alternative method) after equal clotting time (between 45 and 70 min).

d) Storage time: Two SSTs were collected from each patient (N = 31). All pairs of tubes had equal clotting time (between 45 and 120 min), and were centrifuged at 1300*g*. From one of the paired SST tubes, 500 µL serum was frozen at –80°C within 4 h (standard method). The other of the paired tubes was left at room temperature for 48 hours, then 500 µL serum from this tube was frozen at –80°C (alternative method).

e) Different instruments: Aliquots of serum samples from the routine (~ 500) were analyzed, one per day, at equal time points on both Modular I (standard method) and Modular II (alternative method).

Figure 2 presents the description of the standard method, the alternative methods where the maximum deviations for the continuous uncertainty sources are shown, and the modelled current practice. The probability distributions were compatible with actual use in our laboratory.

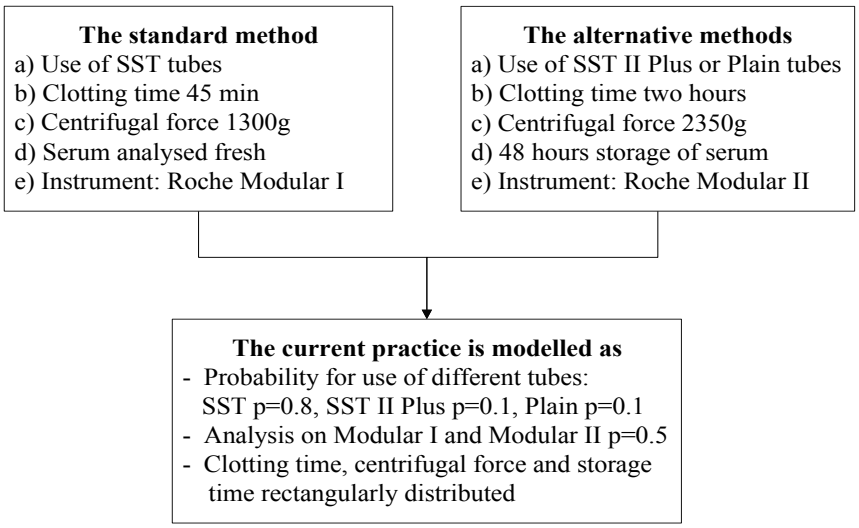

Figure 2. The description of the standard method, the alternative methods, and the modelled current practice (Paper I).

3.4.2 Statistics: Uncertainty budget

Figure 3 shows the design of study I. The mean and variances of differences were calculated between paired observations from comparing a standard method with an alternative method. When comparing the results from the standard method with an alternative method, the results from one of the methods can show a permanent tendency to deviate from the other. We call this systematic effect a bias, and it was estimated as mean of differences. The confidence interval (CI) for the mean of differences could be expected to be much narrower than the CI for the mean values for each data set, because the between-individual variation is not included. It is realistic to presume approximately normal distribution of the differences. Because we were

interested in the variance due to preanalytical treatment, the variance of differences actually measured, was adjusted for analytical variation (66).

We modelled discrete and continuous uncertainty sources separately (Figure 3). If the set of all possible values of a random variable is a countable set, then it is called a discrete random variable (67). For the discrete uncertainty sources, such as using different types of tubes, a number of alternative treatments were defined, together with their probabilities in current practice (Figure 2). The distribution should be estimated from frequency data on the use of each treatment within the laboratory. With increasing probability for use of the standard treatment, the bias and variance will decrease. With increasing probability for use of the alternative treatments, the bias and variance will increase (until a definite limit).

A random variable is called a continuous random variable if it is assumed capable of attaining any value in some interval, and not just discrete points (67). For the continuous source, such as clotting time, we assumed that the treatments in current practice constituted a continuum with the standard treatment at one end, and a maximally distant alternative treatment at the other end. It was assumed that the actual treatment is rectangularly distributed, because the uncertainty variable is within a restricted interval. The rectangular or so-called uniform probability density is zero outside the particular interval, and within the interval, the probability density is a positive constant (30). It was assumed linearity of means and SDs for the continuous source of uncertainty. These assumptions have not been empirically justified.

For each source the expectation and standard deviation of the differences based on the paired data were estimated. The formulas for double expectation and double variance were used to estimate the expectations and variances of the differences for each uncertainty source, assuming that the uncertainty sources varied randomly according to a known distribution (67). It was assumed that the differences are independent of each other, and that each of these differences has the same distribution as if the actual uncertainty source is compared with a fully standardized situation. Under these

conditions, the expected difference D (combined bias) is equal to the sum of the expected individual differences for each uncertainty source:

$$ED = ED_1 + ED_2 + ED_3 + \ldots + ED_k$$

Under the same conditions, the variance of the combined differences is equal to the sum of the expected individual variances:

$$Var(D) = Var(D_1) + Var(D_2) + Var(D_3) + \ldots + Var(D_k)$$

In order to alleviate the burden of computations, Microsoft Excel sheets were prepared and published online as "Supplemental Data – Uncertainty calculations" to Paper I. An example of the calculations sheets is included together with Paper I. An appendix, containing assumptions and modelling details, was published as "Supplemental Data – Appendix" to Paper I. The appendix is included together with Paper I.

The biases and SDs estimated in study I do not refer to the same biases and SDs as in study II and III.

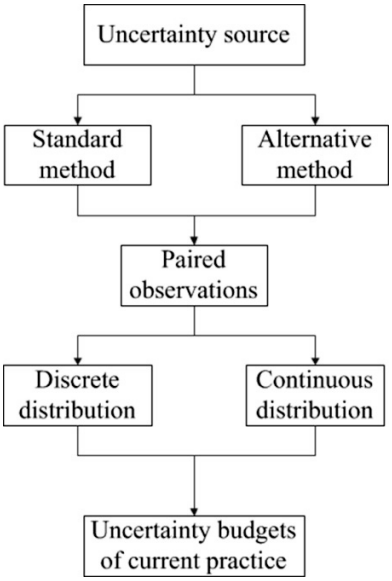

Figure 3. The design of study I.

3.4.3 Methodological considerations

When comparing the results from the standard method with an alternative method, the estimates of the mean and the SD of differences are meaningful only if we can assume that the bias and variation is uniform for the whole concentration range of the test results. The usual deviation from this assumption is an increase or decrease in the variability of the differences (plotted along the y-axis) with increasing concentration at the x-axis (66), where the mean of the test results from both methods are plotted. The assumptions were checked graphically by evaluating the Bland Altman plots, and were evaluated to be acceptable.

We assumed that a rectangular distribution fitted the continuous uncertainty source. A triangular distribution may also be suitable as a continuous distribution for the clotting

time, centrifugal force or storage time, but since we did not have any specific information about the distribution, we chose the rectangular distribution. A simplifying approximation included linearity of the means and SDs for continuous sources of uncertainty. Specifically, they are linear functions of the treatment actually used with slopes α and β, respectively, as detailed in Paper 1.

Current practice for the categorically distributed uncertainty sources was modelled in accordance with our practice. This demands that the probabilities of using the different kind of tubes have to be known and constant, and that one tube is defined as the "gold standard" tube. The bias between methods at two instruments was included in the uncertainty budget. However, a stable bias between the instruments was not verified.

3.5 Paper II and III

3.5.1 Specimen handling

In study II, the phlebotomy and the sample handling were performed optimally according to existing standards (6;13). In order to estimate the uncertainty caused by the phlebotomy, venipunctures were performed on both arms of each participant by using 21-gauge straight green needles. Two SST and two RST tubes were collected in random order from each arm, for a total of 8 tubes from each person. The tubes were completely filled, mixed gently by 5 inversions, and put in a vertical position. Table 3 presents the optimal practice for study II, the optimal and alternative practice in study III, and the model estimates in study II and III.

In study III, experiment 1, four SST gel tubes were collected from each arm by using 21-gauge straight green needles. Two SST tubes, randomly chosen from each arm, i.e. a total of 4 tubes from each patient, were transported after 10 min clotting time by the pneumatic tube system (TranspoNet Pneumatic Tube Systems, Swisslog, Switzerland) installed between the intensive care unit and LKB. Blood samples were padded in bubble plastic before transportation, to avoid the samples to toss around inside the

cartridge. The other 4 gel tubes were manually delivered to the laboratory. The average duration of both pneumatic and hand delivery transport was approximately 2.5 min.

In study III, experiment 2, four SST tubes were collected from one arm using a 21-gauge straight green needle, and four SST tubes from the other arm using a 23-gauge butterfly needle (Figure 4). Two SST tubes, randomly chosen from each arm, i.e. a total of 4 tubes from each patient, were optimally mixed by gently inverting the tubes 5 - 6 times immediately after the phlebotomy. The other 4 gel tubes were mixed by only one inversion.

In experiment 1 and 2, the clotting time was 30 min, the tubes were centrifuged for 10 min at 1600 g, and the specimens were analysed in duplicates within 4 hours after the phlebotomy.

Table 3. The design and model estimates in Paper II and III.

Paper II	**Optimal practice**		**Estimated SDs**	**Estimated biases**
	RST tubes	SST II Advance tubes	Between-venipuncture[a]	RST vs SST
	5 min clotting time	30 min clotting time	Preanalytical (excluding the between-venipuncture)[b]	Left arm vs right arm
	Centrifugal force 1600 g	Centrifugal force 1600 g	Measurement repeatability[a]	
	Analysed on average 33 min (range 21-62 min) after phlebotomy	Analysed on average 61 min (range 46-95 min) after phlebotomy		
Paper III	**Alternative practice**	**Optimal practice**	**Estimated SDs**	**Estimated biases**
Experiment 1	Manual delivery	Pneumatic tube	Between-venipuncture[a]	Pneumatic vs manual delivery
			Preanalytical (excluding the between-venipuncture)[b]	
			Measurement repeatability[a]	
Experiment 2	Green needles	Butterfly needles	Between-venipuncture[a]	Butterfly vs green needles
	Optimal mixing	Suboptimal mixing	Preanalytical (excluding the between-venipuncture)[b]	Suboptimal vs optimal mixing
			Measurement repeatability[a]	

[a] Estimated using the whole data set.
[b] Estimated separately for alternative practice and optimal practice, and for the whole data set.

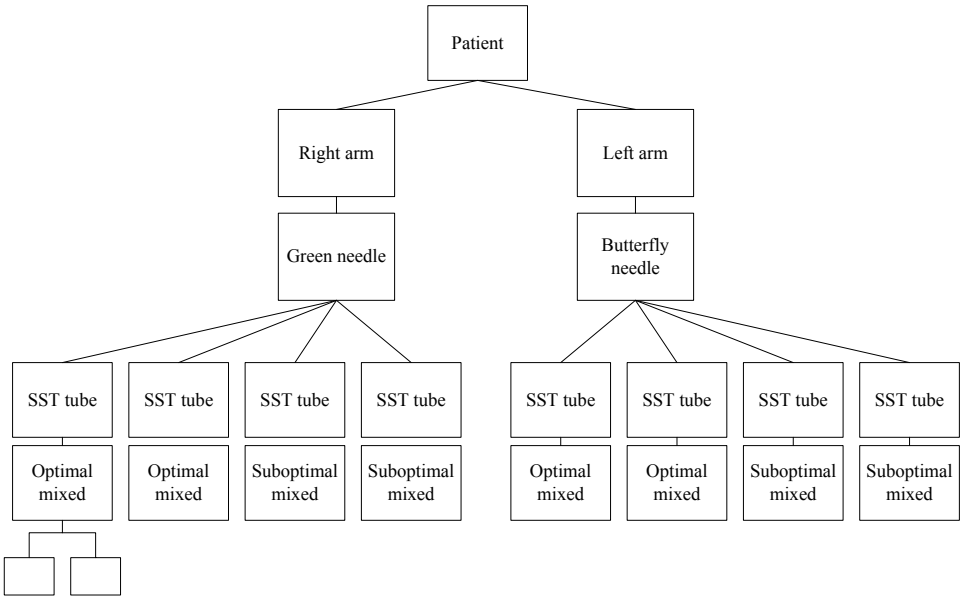

Figure 4. In study III, experiment 2, blood was collected into 4 SST tubes from each arm, by using green straight needle in one arm, and butterfly needle in the other. The blood was mixed differently, and the specimens were analysed in duplicates.

3.5.2 Statistics: Linear mixed effects models

The data in study II and III were analyzed by use of linear mixed-effects models (46). The statistical calculations were done in R (The R Foundation for Statistical Computing, Wien, Austria) with the package nmle (Linear and Nonlinear Mixed Effects Models) (68). The level for statistical significance was set to 0.05, and 95% CIs for both the fixed and random effects were calculated. Comparisons of the SDs for random effects were performed by evaluation of the overlap of their CIs, and the SDs were considered significantly different when their CIs did not overlap. Mixed-effects

models allow analyses of multi-level data, and thereby allow separate estimates of fixed and random effects. Multi-level data can be achieved by collecting and analyzing the blood samples from several individuals (individual level), performing two venipunctures on each individual (venipuncturing level), collecting several tubes from each venipuncture (tube level), and by analyzing each blood tube in duplicate (measurement level).

In study II and III, random effects are expressed as standard deviations for variation between groups at each level, e.g. between individuals (the between-individual biological variation), between arms (the between-venipuncture variation), between tubes from each arm (the preanalytical variation), and duplicates of each tube (measurement repeatability). The random variations are assumed to be independent of each other. The between-individual SDs were not presented in the papers, because the estimates were not relevant in our research. Only by also estimating the measurement repeatability the preanalytical variation will be correctly estimated, and the preanalytical and analytical variation could be compared to each other. Restricted maximum likelihood is the most common technical method used in linear mixed-effects models to estimate the standard deviations.

Figure 5 presents the design of study II. The minimal inevitable preanalytical uncertainty will influence all test results, and may be compared with intra-individual biological variation and analytical variation. The collection of blood into both RST and SST tubes enabled us to detect any differences in preanalytical SDs and fixed effects, between the two clusters of data.

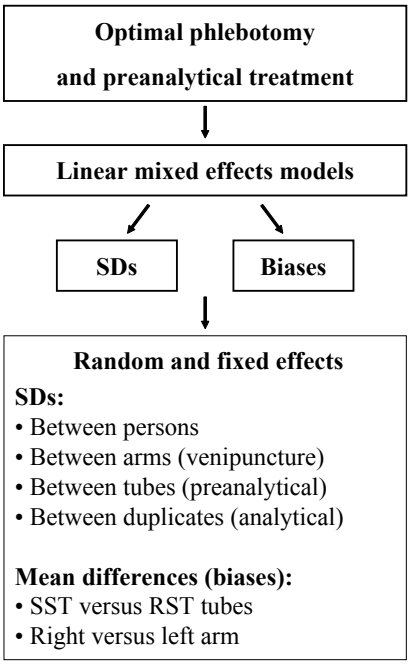

Figure 5. The design of the model in Paper II.

In study III, the preanalytical SDs (excluding the between venipuncture SDs) were estimated separately for each of the paired treatments. The paired treatments involved samples that were suboptimally treated compared with optimally treated:

- transportation in pneumatic tube vs with manual delivery (experiment 1)
- collecting blood using butterfly needles vs green needles (experiment 2)
- mixing blood tubes by only one inversion vs optimally mixed (experiment 2)

We hypothesized that the suboptimal treatment causes a higher preanalytical SD and a bias in test results compared with the optimally treated samples. Biases are connected

to explanatory variables and not to the levels, e.g. type of tubes or needles, or different transportation.

A statistical model for the empirical estimation of uncertainty (18) can be compared with our studies:

$$s^2_{total} = s^2_{between\text{-}target} + s^2_{sampling} + s^2_{analytical}$$

The between-target SD represents the variation between targets, and may be compared with the between-venipuncture SD. The sampling SD is the between-sample SD on one target, and may be compared with the between-tube or preanalytical SD (excluding between-venipuncuture SD). The analytical SD is the between-analysis variation on one sample, and may be compared with the measurement repeatability SD.

3.5.3 Methodological considerations

The widths of the CIs

Occasionally, in linear mixed-effects models, the estimation of random effects results in very wide 95% CI, indicating that the model is unstable. The non-convergent or extremely wide intervals for some analytes presumably result from a model that is too large for the size of the data set, given the actual variation in the analytes in question. The random structure of the model should be simplified when the estimates are close to zero, by respecifying the model to include fewer random effects (68).

The higher number of participants, the narrower the width of the CIs will be both for the fixed and random effects, but the variability of the test results will also influence the width of the CIs. The CIs of the between-venipuncture SDs are pretty wide, compared with the CIs of preanalytical SDs and the measurement repeatability in study II (Figure 6).

Skewed data

Both in experiment 1 and 2 in study III, the distributions of test results were particularly positively skewed for some of the analytes, generally with high test results from only one patient out in the right tail. Test results from blood collections on healthy volunteers in study II, did not have any high values in the tail, and consequently the test results were less skewed than the test results from blood collected on patients in the out-patient clinic in study III.

In study III, the distributions of the test results in the underlying population were evaluated. The number of test results from the patient data system varied between several hundred to thousands for the different analytes. The test results from the patient data system had more or less skewed distributions, and the test results in the selected samples of patients in experiment 1 and experiment 2, would probably follow the same distribution. The test results for the skewed analytes in study III were then log transformed, in order to get them more normally distributed, but the log transformation caused overcompensation for some analytes, giving a left tailed instead of a right tailed skewed distribution. To recalculate the log transformed estimates of random and fixed effects back to ordinary data is difficult, and because also the log transformed estimates are difficult to interpret, we chose not to use log transformed data. The main analysis was performed using all the data, but sensitivity analysis was done by excluding the high test results far out in the right tail of the distribution. Generally, the high results should not be removed from the main analysis, because they are part of the variables' natural variation.

3.6 Combining bias and uncertainty, a follow-up study

3.6.1 Mean square error budget

Mean squared error (MSE) is defined as the expected mean squared deviation from a reference value, or a true value, and covers both the bias and the variance of the differences (67). The combined bias and SD from the uncertainty budget (Paper I) are combined to a mean square error defined as:

$MSE = bias^2 + SD^2$

Root mean square error (RMSE) is defined as $\sqrt{MSE}$.

The preanalytical uncertainty interval can then be calculated as:

Measured value $\pm 2 \cdot$ RMSE

RMSE has the same units as the quantity being estimated.

The MSE budget is presented under Results, and was not presented in Paper I.

3.6.2 Recalculation of bias to random variation

The fixed effects estimated in Paper II and III may be a basis for the calculation of random variation, by connecting the fixed effects to assumptions which can be empirically stated, built on data about clinical practice, or some "what if" assumptions totally or partly built on clinical opinion. An example on how a bias can be recalculated to random variation is presented under Results. The recalculated random variation can be compared with the random effects estimated directly in the model. These calculations have not been presented in any of our papers, and is planned to be the subject of further work.

4. RESULTS

Table 4 presents an overview of the results separately for Paper I, II and III, and the analytes examined in each study.

Table 4. An overview of the results separately for Paper I, II and III, including the analytes examined.

Analytes	**Results Paper I**
Glucose Calcium Magnesium Creatinine	**The combined expected biases (SD) in the preanalytical uncertainty budgets:** *Calcium:* –0.011 (0.0182) mmol/L *Creatinine:* 0.5 (1.81) µmol/L *Glucose:* –0.15 (0.130) mmol/L[a] *Magnesium:* 0.006 (0.026) mmol/L The uncertainty sources contributing to the uncertainty budgets were using different blood tubes and instruments, prolonged clotting time, centrifugal force, and storage time.
	Results Paper II
Albumin, ALP, ALT, Calcium, Cholesterol, CK, Creatinine, GGT, Glucose, HDL-C, H-index, LD, Magnesium, Potassium, Sodium, Triglycerides	**Fixed effects** Statistical significant mean differences ($p<0.05$) were seen between SST vs RST tubes for: *Albumin, calcium, cholesterol, glucose[a], H-index, LD, magnesium, and potassium* **Random effects (results from both the SSTs and the RSTs are included)** *LD:* Preanalytical SD (excluding between-venipuncture) was significantly higher than measurement repeatability. *Potassium:* Preanalytical SD and the between-venipuncture SD were significantly higher than the measurement repeatability. *Glucose:* Between-venipuncture SD was significantly higher than the preanalytical SD and the measurement repeatability.
	Results Paper III
Albumin, ALP, ALT, Bilirubin, Calcium, Cholesterol, CK, Creatinine, Folate, GGT, Glucose, HDL-C, H-index, Iron, LD, Magnesium, Phosphate, Potassium, Sodium, Total protein, Triglycerides, Uric acid	**Fixed effects** Statistical significant mean differences ($p<0.05$) were seen between: Pneumatic tube transport vs manual delivery for: *LD and magnesium* Using butterfly needles vs green needles for: *Calcium, CK and LD* Suboptimal vs optimal mixing for: *Iron* **Random effects** *CK and glucose:* Preanalytical SD (excluding the between-venipuncture SD) for samples transported in the pneumatic tube system was significantly higher than the manually delivered samples. *ALP:* Preanalytical SD (excluding the between-venipuncture SD) for samples collected using butterfly needles was significantly higher than the samples collected by using green needles.

[a] Glucose was the only analyte falling outside the quality specifications for analytical bias, when the significant biases between different preanalytical treatments in Paper II and III, and the combined biases from the uncertainty budget in Paper I, were compared with quality specifications.

4.1 Paper I

4.1.1 Main results

The expected individual biases and variances, between current practice and the standard method for each uncertainty source, were summed up, in order to estimate the combined expected bias and variance. The combined expected biases in the uncertainty budgets were for calcium –0.011 mmol/L, creatinine 0.5 µmol/L, glucose –0.15 mmol/L and magnesium 0.006 mmol/L (Table 2, Paper I).

For glucose, the uncertainty budget shows that prolonged clotting time (-0.091 mmol/L) and storage of serum on gel up to 48 hours at room temperature (-0.058 mmol/L) gave the greatest contribution to the combined bias (Table 2, Paper I). Using different kinds of tubes and instruments, and prolonged centrifugal force gave minor contributions. For calcium, prolonged clotting time (-0.010 mmol/L) gave the greatest contribution to the combined bias, and for creatinine, the storage of serum on gel up to 48 hours (1.1 µmol/L). For magnesium it was the mean difference at 0.014 mmol/L between the SST II Plus vs SST that gave the greatest contribution from the paired data (Table 1, Paper I). But in the uncertainty budget the contribution from using different blood tubes was small, because the probability for using the SST II Plus tube was modelled to P=0.1, using Plain tubes to P=0.1, and using SST tubes to P=0.8.

4.2 Paper II and III

4.2.1 Main results, Paper II

The uncertainty chain starts with the choice of tube. Statistically significant mean differences were seen between SST tubes vs RST tubes for 7 of the 15 analytes (Table 1, Paper II): Calcium 0.013 mmol/L, glucose -0.16 mmol/L, magnesium 0.005 mmol/L, and H-index 1.54 ($P<0.001$), and for albumin 0.16 g/L, cholesterol 0.025 mmol/L, LD 1.8 U/L and potassium 0.039 mmol/L ($P<0.05$). There were no

significant fixed effects indicating differences between venipunctures in left vs the right arm.

The preanalytical SDs (excluding between-venipuncture) estimated separately for use of the SST tubes and RST tubes were not found to be significantly different from each other (Table 2, Paper II), therefore, the results from both tubes were included in the subsequent calculations of random variation. For glucose, the between-venipuncture SD at 0.20 mmol/L, turned out to be the dominant source of variation (Table 3, Paper II). For LD and potassium, the preanalytical SDs (excluding between-venipuncture SD) were significantly higher than the measurement repeatability SDs (Table 3, Paper II). The increase of the total uncertainty is more than 100% for albumin, ALP, CK, glucose, LD, and potassium when the preanalytical contribution is added to the measurement repeatability variation (Table 5). For ALT, sodium and magnesium, the combined CV% is only slightly larger than the measurement repeatability, meaning that the preanalytical variation is negligible. The CV% in Table 5 is calculated based on the mean of the test results and SD estimates from study II (Table 3, Paper II). The CV% can be combined as long as they share a common mean (5).

Table 5. The table presents the preanalytical variation including the between-venipuncture variation, the measurement repeatability, and the preanalytical and measurement repeatability combined for each analyte in study II.

Analytes	Mean (Range)[a]	CV%		
		Preanalytical variation including between venipuncture[b]	Measurement repeatability	Preanalytical and measurement repeatability combined[c]
Albumin, g/L	46.1 (41.6 – 50.4)	1.3	0.7	1.5
ALP, U/L	71.1 (46 – 107)	1.5	0.7	1.7
ALT, U/L	27.8 (10 – 67)	2.1	4.2	4.6
Calcium, mmol/L	2.40 (2.20 – 2.53)	0.7	0.6	1.0
Cholesterol, mmol/L	5.23 (3.57 – 7.65)	1.6	1.0	2.0
CK, U/L	118.3 (52 – 191)	1.5	0.8	1.7
Creatinine, µmol/L	70.4 (57 – 92)	1.5	1.0	1.8
GGT, U/L	26.2 (8 – 160)	1.9	2.3	3.0
Glucose, mmol/L	5.2 (4.04 – 8.50)	4.0	1.1	4.2
HDL-C, mmol/L	1.82 (1.03 – 3.07)	1.4	1.0	1.8
LD, U/L	177.3 (97 – 223)	2.3	1.1	2.5
Magnesium, mmol/L	0.85 (0.75 – 0.94)	0.7	1.3	1.4
Potassium, mmol/L	4.38 (3.75 – 4.93)	2.7	0.7	2.8
Sodium, mmol/L	140.5 (135.5 – 143.8)	0.2	0.5	0.6
Triglycerides, mmol/L	0.91 (0.49 – 2.18)	2.2	1.5	2.6

[a] The mean and the range of observations in study II.
[b] The preanalytical variation (excluding the between-venipuncture variation) and the between-venipuncture variation combined.
[c] The between-venipuncture and the preanalytical (excluding between-venipuncture) variation, and the measurement repeatability are combined.

4.2.2 Main results, Paper III

The mean differences were estimated in the paired treatments between transporting samples in a pneumatic tube system vs manual delivery, use of butterfly vs green needles, and suboptimal vs optimal mixing (Table 1, Paper III). In experiment 1, transporting samples by the pneumatic tube system added a significant bias to the test results for LD at 4.5 U/L ($P<0.001$) and magnesium at 0.0021 mmol/L ($P=0.003$) (Table 1, Paper III). The H-index was not affected. Statistical significant mean differences were also seen between using butterfly needles vs green needles for calcium and CK, and between suboptimal vs optimal mixing for iron (Table 1, Paper III). LD was significantly lower (-1.6 U/L, $P=0.047$) when using butterfly needles compared with using green needles, when samples with H-index > 40 were excluded.

The preanalytical SDs (excluding the between-venipuncture SD) were estimated separately for each of the paired treatments (Table 2, Paper III). For glucose the preanalytical SD (0.12 mmol/L, 95% CI 0.097 – 0.14) for samples transported in the pneumatic tube system was significantly higher than the preanalytical SD for samples manually delivered (0.077 mmol/L, 95% CI 0.062 – 0.095). The preanalytical SDs (excluding the between-venipuncture SD) estimated based on the whole data set for experiment 1, and correspondingly for experiment 2, are presented together with the preanalytical SDs estimated in study II in Table 6. Correspondingly, the between-venipuncture SDs are presented together in Table 7.

Sensitivity analysis was done by excluding 8 samples with H-index > 40 for the analytes in experiment 2. The remaining samples had H-index ≤ 14. The 8 samples represented duplicates of 4 tubes collected from two venipunctures in two patients. Only the model estimates of LD were substantially influenced by excluding the high H-index samples (Table 6 and 7).

Both in experiment 1 and 2, the distributions of test results were particularly positively skewed for the analytes ALP, ALT, creatinine, GGT, LD, and triglycerides, generally

with high test results from only one of the patients in the right tail. The sensitivity analysis showed that the model estimates were pretty robust against skewed data except for GGT (Table 6 and 7). Generally the CIs were somewhat broader when keeping the high values, but this only slightly affected the model estimates. Some measurement repeatability SDs were on the other hand reduced by excluding the high values (Table 3, Paper III).

4.2.3 Preanalytical SDs

Table 6 presents the preanalytical SDs (excluding between-venipuncture SD) (95% CI) estimated in study II, and in experiment 1 and 2 in study III. The CIs of the preanalytical SDs overlapped for 17 of 21 analytes. For creatinine, glucose and LD, the preanalytical SDs were significantly lower for the optimally treated samples (study II) compared with one or both of the estimates from study III. But when excluding samples with H-index>40 for LD (experiment 2, Paper III), the preanalytical SD in experiment 2 was similar to the preanalytical SD estimated for the optimally handled samples. Calcium was the only analyte showing higher preanalytical SD for the optimally treated samples.

Table 6. Comparison of the preanalytical SDs (excluding between-venipuncture SD) estimated in study II and III.

Analytes	Preanalytical SDs (excluding between-venipuncture SD) (95% CI)		
	Study II[a]	**Study III**[b] **(experiment 1)**	**Study III**[c] **(experiment 2)**
Albumin, g/L	0.38 (0.31 – 0.45)	0.28 (0.23–0.33)	0.31 (0.26–0.36)
ALP, U/L	0.64 (0.54 – 0.76)	0.37 (0.23–0.59)	0.52 (0.44–0.60)
ALT, U/L	0.32 (0.094 – 1.076)	0.24 (0.039–1.42)	0.19 (0.029–1.20)[d]
Bilirubin, µmol/L	Not analysed	0.069 (0.040–0.12)	0.086 (0.047–0.16)
Calcium, mmol/L	0.012 (0.010 - 0.016)	0.0071 (0.0053–0.0095)	0.0084 (0.0068–0.010)
Cholesterol, mmol/L	0.048 (0.039 – 0.060)	0.030 (0.022–0.041)	0.031 (0.024–0.040)
CK, U/L	1.02 (0.85 – 1.23)	0.89 (0.75–1.07)	1.13 (0.98–1.30)
Creatinine, µmol/L	0.58 (0.46 – 0.72)	0.66 (0.55–0.80)	0.94 (0.80–1.10)
Folate, mmol/L	Not analysed	0.17 (0.052–0.58)	0.28 (0.17–0.45)
GGT, U/L	0.24 (0.13 – 0.46)	0.50 (0.098–2.55) 0.24 (0.18–0.32)[e]	0.52 (0.44–0.62) 0.27 (0.20–0.36)[f]
Glucose, mmol/L	0.07 (0.06 – 0.08)	0.097 (0.086–0.11)	0.10 (0.087–0.11)
HDL-C, mmol/L	0.014 (0.010 – 0.018)	0.011 (0.0050–0.024)	0.0094 (0.0079–0.011)
Iron, µmol/L	Not analysed	0.11 (0.090–0.13)	0.14 (0.12–0.17)
LD, U/L	3.2 (2.8 – 3.7)	5.2 (4.6–5.8)	4.4 (3.9–4.9) 2.5 (2.2–2.8)[g]
Magnesium, mmol/L	0.004 (0.002 – 0.008)	0.0011 (0.000036–0.0082)	0.0021 (0.0010–0.0044)
Phosphate, mmol/L	Not analysed	0.0059 (0.0046–0.0075)	0.0069 (0.0057–0.0082)
Potassium, mmol/L	0.092 (0.080 – 0.11)	0.072 (0.065–0.081)	0.080 (0.072–0.090)
Sodium, mmol/L	0.26 (0.14 – 0.47)[d]	Instability[h]	0.16 (0.082–0.32)
Total protein, g/L	Not analysed	0.40 (0.32–0.50)	0.52 (0.45–0.60)
Triglycerides, mmol/L	0.011 (0.010 – 0.014)	0.0087 (0.0057–0.013)	0.013 (0.011–0.015)
Uric acid, µmol/L	Not analysed	0.33 (0.049–2.25)	0.63 (0.42–0.95)

[a] Optimally treated SST and RST tubes (whole data set).
[b] Samples transported by pneumatic tube system and manual delivery (whole data set, experiment 1).
[c] Blood collected with butterfly and green needles, and optimally and sub-optimally mixed (whole data set, experiment 2).
[d] Because of wide 95% CIs, the between-venipuncture and preanalytical SD were combined to a common estimate.
[e] High values of about 1495 U/L from one patients, were excluded.
[f] High values of about 743 U/L from one of the patients, were excluded.
[g] 4 samples with H-index > 40 excluded in addition to the already 4 samples automatically excluded at the instrument.
[h] Wide 95% CIs, and simplifying the model also caused instability.

4.2.4 Between-venipuncture SDs

Table 7 presents the between-venipuncture SDs (95% CI) estimated in study II, and in experiment 1 and 2 in study III. For all of the analytes, except for glucose, the CIs of the between-venipuncture SDs were overlapping. For glucose, the between-venipuncture SD in experiment 1 at 0.32 mmol/L (green needles were used in both arms), is significantly higher than the between-venipuncture SD in experiment 2 at 0.15 mmol/L, where green needle was used in one arm and a butterfly needle in the other. For LD, the between-venipuncture SDs were overlapping when the high H-index samples were excluded.

Table 7. Comparison of the between-venipuncture SDs estimated in study II and III.

Analytes	Between-venipuncture SDs (95% CI)		
	Study II[a]	**Study III**[b] **(experiment 1)**	**Study III**[c] **(experiment 2)**
Albumin, g/L	0.47 (0.32 – 0.69)	0.48 (0.36–0.64)	0.38 (0.27–0.53)
ALP, U/L	0.87 (0.59 – 1.27)	0.68 (0.49–0.95)	0.70 (0.51–0.95)
ALT, U/L	0.47 (0.26 – 0.86)	0.17 (0.020–1.4)	0.19 (0.029–1.20)[d]
Bilirubin, µmol/L	Not analysed	0.14 (0.10–0.20)	0.10 (0.061–0.17)
Calcium, mmol/L	0.012 (0.010 - 0.019)	0.010 (0.0074–0.014)	0.0099 (0.0069–0.014)
Cholesterol, mmol/L	0.071 (0.049 – 0.104)	0.050 (0.037–0.069)	0.058 (0.043–0.079)
CK, U/L	1.51 (1.04 – 2.19)	0.92 (0.65–1.30)	1.09 (0.75–1.58)
Creatinine, µmol/L	0.90 (0.62 – 1.31)	1.49 (1.12–1.97)	1.61 (1.20–2.16)
Folate, mmol/L	Not analysed	0.11 (0.019–0.67)	0.14 (0.036–0.56)
GGT, U/L	0.42 (0.28 – 0.65)	0.83 (0.47–1.47) 0.24 (0.15–0.37)[e]	0.69 (0.50–0.95) 0.26 (0.16–0.41)[f]
Glucose, mmol/L	0.20 (0.14 – 0.27)	0.32 (0.24–0.41)	0.15 (0.11–0.21)
HDL-C, mmol/L	0.022 (0.015 – 0.032)	0.017 (0.011–0.026)	0.014 (0.010–0.019)
Iron, µmol/L	Not analysed	0.16 (0.12–0.22)	0.18 (0.13–0.26)
LD, U/L	2.4 (1.5 – 3.9)	2.1 (1.1–4.2)	4.8 (3.5–6.6) 2.7 (1.9–3.7)[g]
Magnesium, mmol/L	0.004 (0.002 – 0.008)	0.0049 (0.0035–0.0068)	0.0047 (0.0033–0.0065)
Phosphate, mmol/L	Not analysed	0.0080 (0.0058–0.011)	0.012 (0.0090–0.016)
Potassium, mmol/L	0.075 (0.048 – 0.12)	0.094 (0.070–0.13)	0.067 (0.046–0.096)
Sodium, mmol/L	0.26 (0.14 – 0.47)[d]	Instability[h]	0.21 (0.13–0.34)
Total protein, g/L	Not analysed	0.85 (0.63–1.13)	0.66 (0.48–0.90)
Triglycerides, mmol/L	0.016 (0.011 – 0.024)	0.013 (0.0088–0.018)	0.018 (0.013–0.025)
Uric acid, µmol/L	Not analysed	1.13 (0.81–1.59)	0.76 (0.50–1.15)

[a] Optimally treated SST and RST tubes. Green needles were used in both arms.
[b] Samples transported by pneumatic tube system and manual delivery. Green needles were used in both arms.
[c] Tubes collected with butterfly and green needles, and optimally and sub-optimally mixed blood samples.
[d] Because of wide 95% CIs, the between-venipuncture and preanalytical SD were combined.
[e] High values of about 1495 U/L from one patients, were excluded.
[f] High values of about 743 U/L from one of the patients, were excluded.
[g] 4 samples with H-index > 40 excluded in addition to the already 4 automatically excluded at the instrument.
[h] Wide 95% CIs, and simplifying the model also caused instability.

4.2.5 The width of confidence intervals

The width of the CIs for the random effects in study II

Figure 6 presents the between-venipuncture SD, the preanalytical SD, and the measurement repeatability SD including the 95% CIs estimated in study II. The figure shows that the measurement repeatability SDs had the narrowest CIs, and the between-venipuncture SDs the broadest CIs for 11 of the 14 analytes.

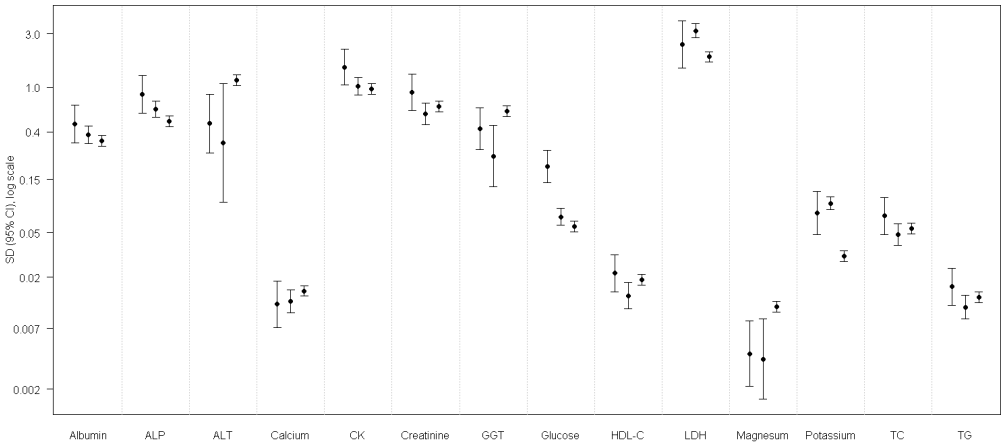

Figure 6. The figure presents, in respective order, the between-venipuncture SD, the preanalytical SD, and the measurement repeatability SD including 95% CIs estimated in study II. The vertical axis displays the size of the SDs (log scale).

Comparison of the width of the CIs for preanalytical SDs in study II and III

Table 8 presents some examples on the relative width of the 95% CIs of the preanalytical SDs (excluding between-venipuncture SD) in study II and III. The relative width of the CIs is calculated as (the data from albumin in study II is used as example): (0.45–0.31) / 0.38 · 100 % = 37%

Table 8. Relative width of the 95% CIs for the preanalytical SDs (excluding between-venipuncture SD) in study II and III for five analytes.

	Preanalytical SDs (excluding between-venipuncture) (95% CI) Relative width of the CIs		
Analytes	**Study II[a]**	**Study III[b] (experiment 1)**	**Study III[b] (experiment 2)**
Albumin, g/L	0.38 (0.31– 0.45)	0.28 (0.23–0.33)	0.31 (0.26–0.36)
Relative width	37%	36%	32%
Creatinine, µmol/L	0.58 (0.46 – 0.72)	0.66 (0.55–0.80)	0.94 (0.80–1.10)
Relative width	45%	38%	32%
Glucose, mmol/L	0.07 (0.06–0.08)	0.097 (0.086–0.11)	0.10 (0.087–0.11)
Relative width	29%	25%	23%
LD, U/L	3.2 (2.8– 3.7)	5.2 (4.6–5.8)	4.4 (3.9–4.9) 2.5 (2.2–2.8)[a]
Relative width	28%	23%	23 % 24%[c]
Potassium	0.092 (0.080 – 0.11)	0.072 (0.065–0.081)	0.080 (0.072–0.090)
Relative width	33%	22%	23%

[a] Study II included n=20, healthy volunteers
[b] Study III included n=30 patients in experiment 1 and correspondingly in experiment 2
[c] 4 samples with H-index > 40 excluded in addition to the already 4 automatically excluded at the instrument.

4.3 Biases compared with quality specifications

In order to evaluate the size of the biases in the studies, the combined biases from the uncertainty budgets (study I), and the statistically significant biases (95% CI) from study II and III, are compared with the desirable quality specifications for analytical bias derived from biological variation (39) (Table 9). All biases, except for glucose, were within the quality specifications. For glucose, the combined bias in the uncertainty budget at -0.15 mmol/L, and the bias between SST vs RST at -0.16 mmol/L, was falling outside the quality specification at 0.11 mmol/L.

Table 9. Comparison of the combined biases from the uncertainty budgets (study I), and the statistically significant biases (95% CI) in study II and III with desirable quality specifications for analytical bias derived from biological variation.

Analytes	Different comparisons	Mean[a]	Biases (95% CI)	Qual. spec. %[b]	Qual-spec. abs. value[c]
Albumin, g/L	SST vs RST	46.1	0.16 (0.016 to 0.29)	1.3	0.60
Calcium,mmol/L	Uncertainty budget	2.40	-0.011	0.8	0.019
	SST vs RST	2.40	0.013 (0.0079 to 0.018)	0.8	0.019
	Butterfly vs green	2.33	-0.0072 (-0.012 to -0.0012)	0.8	0.019
CK, U/L	Butterfly vs green	107	-0.75 (-1.42 to -0.075)	11.5	12.3
Cholesterol,mmol/L	SST vs RST	5.23	0.025 (0.0060 to 0.045)	4.0	0.21
Creatinine, µmol/L	Uncertainty budget	100[d]	0.5	4.0	4.0
Glucose, mmol/L	Uncertainty budget	5.2	-0.15	2.2	0.11
	SST vs RST	5.2	-0.16 (-0.18 to -0.13)	2.2	0.11
Iron, µmol/L	Suboptimal vs optimal mixing[g]	16.9	0.065 (0.016 to 0.11)	8.8	1.49
LD, U/L	SST vs RST	177.3	1.8 (0.8 to 2.9)	4.3	7.6
	Pneumatic vs manual	210	4.5 (3.1 to 5.8)	4.3	9.0
	Butterfly vs green	186	-1.6 (-3.2 to -0.021)	4.3	8.0
Magnesium, mmol/L	Uncertainty budget	0.85	0.006	1.8	0.015
	SST vs RST	0.85	0.005 (0.003 to 0.008)	1.8	0.015
	Pneumatic vs manual	0.80	0.0021 (0.00074 to 0.0035)	1.8	0.014
Potassium, mmol/L	SST vs RST	4.38	0.039 (0.010 to 0.068)	1.8	0.079

[a] Mean of the test results in each study. Uncertainty budgets: The mean from one of the other projects are used.
[b] Desirable quality specifications (%) for analytical bias (39).
[c] The quality specifications are calculated from % to absolute values based on the mean in each study.
[d] For creatinine, 100 µmol/L was used to calculate the limit of quality specification.

4.4 Combining bias and uncertainty

4.4.1 Mean squared error budget

Table 10 presents the calculation of the mean squared error (MSE) based on the combined bias and variance from the uncertainty budget (Table 2, Paper I).

$MSE = bias^2 + SD^2$

$RMSE = \sqrt{MSE}$

Measured value ± 2· RMSE give the preanalytical uncertainty interval. The preanalytical uncertainty intervals do not include the analytical SDs.

Table 10. Calculation of the MSE and the preanalytical uncertainty interval, based on the combined bias and variance from the uncertainty budget for each analyte.

Analytes	Combined bias[a]	Combined variance[b]	MSE[c]	RMSE[d]	Preanalytical uncertainty interval[e]
Calcium, mmol/L	-0.011	0.00033	0.000452	0.0213	± 0.04 mmol/L
Creatinine, µmol/L	0.5	3.28	3.53	1.88	± 3.8 µmol/L
Glucose, mmol/L	-0.15	0.017	0.0395	0.20	± 0.4 mmol/L
Magnesium, mmol/L	0.006	0.00066	0.000695	0.026	± 0.05 mmol/L

[a] Combined bias from the uncertainty budget (Table 2, Paper I)
[b] Combined variance from the uncertainty budget (Table 2, Paper I)
[c] MSE is defined as the expected squared deviation from a reference value, or a true value. $MSE = bias^2 + SD^2$
[d] $RMSE = \sqrt{MSE}$
[e] The preanalytical uncertainty interval is ± 2· RMSE

4.4.2 Recalculation of bias to random variation, an example

For glucose, the estimated mean difference between transporting samples with a pneumatic tube system vs manual delivery was 0.025 mmol/L (Table 1, Paper III). As an example, we presume that 42% of the samples are transported by the pneumatic

tube system. The bias at 0.025 mmol/L can be recalculated to random variation based on the probability for transporting the samples by the pneumatic tube system:

Transport type	Pneumatic tube system	Manual delivery
Probability	0.42	0.58
E (A|transport type)	a + 0.025	a

where "a" is determined by the fixed effects for the other covariates.

A = measured concentration of the analyte

Then, by standard formulas for double expectation and variance,

$EA = EE(A|\text{transport type}) = 0.42 \cdot (a + 0.025) + 0.58 \cdot a = \underline{a + 0.42 \cdot 0.025}$

$EA^2 = 0.42 \cdot (a + 0.025)^2 + 0.58 \cdot a^2 = \underline{a^2 + 2 \cdot 0.42 \cdot 0.025 \cdot a + 0.42 \cdot 0.025^2}$

$Var(A) = (EA^2) - (EA)^2 =$

$a^2 + 2 \cdot 0.42 \cdot 0.025 \cdot a + (0.42 \cdot 0.025^2) - (a^2 + 2 \cdot a \cdot 0.42 \cdot 0.025 + 0.42^2 \cdot 0.025^2) =$

$0.42 \cdot 0.025^2 - 0.42^2 \cdot 0.025^2 = 0.025^2 \cdot 0.42\ (1\text{-}0.42) = \underline{0.00015}$

$SD(A) = \sqrt{Var\ (A)} = 0.025 \sqrt{[0.42\ (1\text{-}0.42)]} = 0.012$ mmol/L

Similar recalculations are planned to be elaborated in further work.

5. DISCUSSION

In this thesis and its associated papers we aimed to establish a modelling framework for estimating preanalytical uncertainty. Knowledge of the relative magnitude of uncertainty sources provide possibilities for modifying the measurement systems to improve the quality of results (2). In study I, the "bottom-up" approach, a model for an uncertainty budget was established, based on the biases from paired data between the standard method for handling blood samples and current practice. The model consisted of five subprojects examining one uncertainty source at a time. But the model did not determine the uncertainty in the optimal practice itself.

In study II, the "top-down" approach, the phlebotomy and sample handling were performed optimally according to existing standards, and by using linear mixed-effects models (46), the minimal preanalytical uncertainty in optimal practice was estimated. This minimal preanalytical uncertainty may be used as a standard of reference for evaluation of preanalytical uncertainty in current practice. Linear mixed-effects models constitute a well-developed and suitable method to identify different sources of variation (46). The model allows clustered data, where both random variation and systematic deviations are estimated, and the model assumptions are made clear.

Study III involved further use of linear mixed-effects models, and the number of preanalytical variables was expanded. We wanted to examine whether specific, preanalytical treatments within current practice would increase the preanalytical variation, and bias test results compared with optimal treatment. Literature search did not lead to similar studies using linear mixed-effects models within the preanalytical field, indicating that this thesis include models that are quite new within the field.

In the following, the specific findings in each paper, the models and methodological considerations will be discussed. Further on, the preanalytical (excluding the between-venipuncture) and the between-venipuncture variations estimated in study II and III

are compared to each other, followed by a discussion on the treatment of biases and the concept of traceability in preanalytical phase.

5.1 Paper I

5.1.1 Main results

Paired data are easy to attain, and are often available in minor projects. For all the different preanalytical uncertainty sources in the budget, standard methods were defined. For the continuously distributed uncertainty sources, the alternative methods were defined as the maximal deviations within current practice. E.g. clotting time can attain any value in some interval, and represents a continuously distributed uncertainty source. Using different kinds of blood tubes involves a finite number of alternative methods, and represents a categorically distributed uncertainty source. The uncertainty contribution from each uncertainty source is easily accessible in the budget (Table 2, Paper I), and solutions can be made directly to reduce the uncertainty. Most important is to reduce the uncertainty from any great contributors, but also several minor contributions can be summed up to give a considerable contribution.

Generally, uncertainty budgets are established from existing knowledge, and should predict the combined uncertainty of future test results (69). It is claimed that the uncertainty budgets should be verified, because uncertainty components may have been overlooked, uncertainty may have been incorrectly estimated, or changes may have occurred (69). Because of medical consequenses, the combined bias, from summing up the individual biases, should not fall outside defined quality specifications.

When comparing the combined biases in the uncertainty budget with quality specifications (Table 9), only glucose at -0.15 mmol/L was falling outside the quality specifications for analytical bias (39). Prolonged clotting and storage time were the greatest contributors to the uncertainty budget, and the concentration of glucose

probably decreased because of glycolysis (54). Because of a variety of experimental designs, it may be difficult to interpret information on stability from the literature (70). Nevertheless, a study showed that glucose decreased, and exceeded the combined analytical and clinical acceptability limit, after storing the blood tubes at 32 ^{o}C in 3 hours compared with the values from separation of the sera from the clot within 30 min (70).

5.1.2 Methodological considerations

Detailed knowledge of the nature of the component and of the sample handling is essential both for designing the study, and for calculating the uncertainty. In practice, the alternative methods are used to a variable extent. The probability distributions seem suitable for the uncertainty sources included in the budget, because the countable alternative treatments characterize the categorically distributed uncertainty sources, while the attainment of any value within an interval, characterize the continuously distributed uncertainty sources.

When the probabilities of e.g. using the alternative blood tubes changes, or if the standard blood tube is exchanged to another tube, new calculations of the bias have to be done. At the laboratory there are usually several instruments analysing the same analyte. Continuity of medical care requires that the comparability of test results produced by different instruments is verified periodically (71). It is recommended to define one instrument as the standard instrument, and the other instruments should then be calibrated up to the standard instrument, in order to attain the same analytical accuracy. There still may be a small bias between the instruments, and the bias could occasionally vary. In study I, this bias was a part of the preanalytical uncertainty budget. Because the mean and SD of differences in the paired data between the standard method and the alternative method are fundamental in the calculations, it is important that they are representative for the whole range of test results.

For continuous uncertainty sources, the assumptions of linearity of the means and SDs of differences have not been empirically justified. Because the interval defined for each continuously distributed uncertainty source is not especially broad, and assuming that the bias at the maximal distance is highest, the estimated biases are probably not overestimated. The assumption of independence, when summing up the expected individual differences and variances seems reasonable, because the uncertainty sources are probably not related to each other, and therefore they are assumed not to influence on each other.

5.2 Paper II and III

5.2.1 Main results

Good standardization is a prerequisite for high power and accurate model estimates. The analysing methods have good analytical precision and trueness, ensuring the reliability and the precision of the model estimates. Both study II and III were well standardized in the sense that randomization was carried out at all levels, e.g. what type of needle should be used on what arm, the order of tubes during the collection, the choice of tubes that should be transported in the pneumatic tube system, and during analyses. The phlebotomy time was supervised, the clotting time was held within acceptable limits, and the time after centrifugation until analysis was supervised. By analyzing the H-index in all samples, the preanalytical quality of the samples was supervised. Examining many analytes, as in our studies II and III, is also a strength, and the finding of small significant differences indicates that the number of observations was acceptable. However, there will always be a certain probability of finding significant differences by chance in studies involving many analytes, repeated measurements and statistical tests.

In study II, the optimal handling of the SST tubes included 30 min clotting time, while for the RST tubes the clotting time was 5 min. Due to shorter clotting time for the RSTs, serum was separated from the erythrocytes approximately 25 min before the

SSTs, and this may explain why the H-index was significantly higher in the SST tubes compared with the RST tubes. Hemolysis may also be the reason why potassium, LD and magnesium showed a significantly higher concentration in the SSTs, and glucose a significantly lower concentration, probably because of glycolysis (Table 1, Paper II). Thus, the choice of correct tube is important to minimize the preanalytical uncertainty, even when following an optimal protocol.

Table 5 shows that preanalytical uncertainty including the between-venipuncture variation is a substantial contributor to the total uncertainty of test results, although the blood samples were treated optimally based on current guidelines. The uncertainty increased nearly 300% for glucose, and potassium, when the preanalytical and the between-venipuncture variations were included in the uncertainty calculations, compared to the measurement repeatability. Thus, those variations are shown not to be negligible even for optimally handled samples.

A study, which included the uncertainty from specimen collection, delay in treatment, and transportation, the preanalytical uncertainty for cholesterol was estimated to 0.7%, for albumin 4.5% and potassium 6.6%, and the combined preanalytical and analytical uncertainty was 4.1%, 6.5%, and 6.8%, respectively (27). In our results on preanalytical variation including the between-venipuncture variation for optimally treated samples (Table 5), albumin and potassium were estimated much lower. A study by Fuentes-Arderiu et al. on premetrological variation estimated by using paired data, where the variance of the differences were adjusted for analytical variation, included uncertainty from venipuncturing both arms, use of different phlebotomists, and different clotting time (23). In that study, the premetrological variation was estimated to be 3.2% for glucose, for albumin 1.3%, calcium 0.7%, and potassium 3.1% (23). Compared with our results shown in Table 5, the estimates are very similar. In an additional study, the 95% uncertainty interval of a measurement result on fasting plasma glucose at 5.9 mmol/L was estimated to be 5.1 – 6.6 mmol/L (26). The interval included uncertainty from patient preparation, position, sampling and measurement procedure. In our study II, the combined SD of preanalytical and analytical SD for

optimally handled samples for glucose was 0.22 mmol/L (Table 3, Paper II), i.e. the 95% uncertainty interval is about ± 0.44 mmol/L. It seems reasonable that our estimate is different, both because our uncertainty interval includes fewer uncertainty sources, and because of the difference in methods for calculating the uncertainty. In the study by Kallner et al. plausible intervals were assumed and the uncertainty estimated from rectangular distributions (26).

The results obtained from any model are estimates, not the truth, and estimates may differ depending on underlying assumptions. The design of the studies in the cited papers, and the differences in the preanalytical variations estimated in the different studies, may indicate that there are a need for clear descriptions of what uncertainty sources are included, what statistical methods are used, and what the assumptions are. This is needed, particularly, if estimated preanalytical uncertainty is to be transferable to other laboratories. This work is a response to a need for a standardized statistical method for estimating preanalytical uncertainty.

In study III, transporting blood tubes by a pneumatic tube system resulted in significantly higher values for LD and magnesium (Table 1, Paper III), probably due to hemolysis or leakage through the cell membrane, as the concentration of LD (62) and magnesium (72) are higher in erythrocytes than in serum. However, the higher preanalytical SD for glucose for blood tubes transported in a pneumatic tube system compared with manual delivery (Table 2, Paper III), is difficult to explain. The shaking and speed of the tubes during the pneumatic tube transport, may increase the glycolysis in some tubes. It may be difficult to transfer research results from one pneumatic tube system to another, because of different technical characteristics of the pneumatic tube system, such as speed, length and time of transport.

Because the preanalytical SDs estimated separately for the paired treatments (Table 2, Paper III) showed minor differences for only a few analytes, the random effects presented in Table 3 and 4 in Paper III, are estimated based on the whole data set separately for each experiment. For most analytes, the preanalytical SDs were about

the same in both experiments, indicating that the preanalytical variation is little influenced by the different preanalytical treatments carried out in study III.

When using butterfly needles, LD was significantly lower -1.6 U/L (P=0.047) (samples with H-index > 40 were excluded) compared with using green needles. The opposite might be expected, because using butterfly needles is associated with higher pressure of the blood, which may result in hemolysis (73). On the other hand, using butterfly needles may give a more stable phlebotomy, and a more gentle cut in the vein compared with the green, straight needles. Mixing the blood tubes by only one inversion compared with recommended mixing of 5-6 inversions, did not influence the test results. In general, the statistically significant systematic effects found between suboptimal vs optimal treatment in Paper III were too small to have any clinical impact.

5.2.2 Preanalytical SDs

The preanalytical SD (excluding venipuncture SD) may be caused by variations in filling of the tubes, sample mixing, clotting time, centrifugal force, and extent of hemolysis. It may be expected that the optimally handled samples in study II have less preanalytical SD than the samples in study III, which included both optimally and suboptimally treated samples (Table 6). Lack of significant difference may indicate that the preanalytical SDs based on optimally treated samples, are reasonable estimates of preanalytical variation, independent of the preanalytical treatment shown here. There are several conditions that may influence both the magnitude of the SDs, and the width of the CIs, and it is difficult to explain the difference or the similarity in preanalytical SDs between the studies. The preanalytical treatment, whether healthy individuals or patients are included, number of participants, concentration and distribution of the test results, are factors that may influence the magnitude of the preanalytical SDs.

Of the measured analytes, seventeen of 21 analytes had overlapping preanalytical SDs (Table 6). For calcium, the only analyte showing higher preanalytical SD for the optimally treated samples, the concentration of calcium was somewhat higher in the samples in study II (Table 3, Paper II) than in study III (Table 3 and 4, Paper III), which may influence on the magnitude of the preanalytical SD. The slightly lower preanalytical SD for glucose at optimal treatment in study II may be caused by less variation in the test results, because of a more perspicuous and standardized treatment, than the different treatments in study III. However, the difference in preanalytical SD is too small to have any clinical significance. For LD, it is difficult to compare the preanalytical SDs between the studies, because LD is sensitive for hemolysis, and hemolysis may be caused by the phlebotomy and by the specific preanalytical treatment. Excluding some samples with H-index > 40 in experiment 2 seems realistic, because the remaining samples then had H-index less than 14.

The test results in study III were somewhat more skewed than the test results in study II. More heterogeneous test results may give wider CIs. However, the number of participants was higher in study III than in study II, that may compensate for the skewed data. Thus, there are no indications that the preanalytical SDs are higher in samples from patients than in blood samples collected on healthy individuals. Consequently, estimating the minimal preanalytical uncertainty by analyzing optimally treated blood samples from healthy individuals, may give an unbiased estimate of the general preanalytical variation of blood samples, collected in an outpatient clinic and suboptimally treated as shown here. Transporting the samples by a pneumatic tube system, collecting blood using butterfly needles, and suboptimal mixing, do not substantially influence the size of the preanalytical uncertainty.

Regarding the design of the study, it may influence on the size of the random and fixed effects. Blood was collected in accordance with current practice into four tubes from each arm. The mean venipuncture time for both arms was about 3 min in study II, and 3.5 min in study III which is considered acceptable. If only two blood tubes were collected from each arm, we would probably need a higher number of participants, in

order to achieve similar narrow CIs. Contrary, collection of e.g. six blood tubes from each arm, would be very different from current practice. By letting several medical technicians perform the phlebotomy instead of only one as in study II and III, the estimates of both the preanalytical and the between-venipuncture SDs would probably be higher. Both variations in filling volume of the blood tubes, in mixing of the blood tubes, clotting time, and degree of hemolysis, would be influenced by the medical technician. In contrast, variation in centrifugation time, storage time before analysis, and room temperature, are procedures not dependent on the phlebotomist.

5.2.3 Between-venipuncture SDs

The between-venipuncture variation turned out to be somewhat higher than the preanalytical variation (excluding the between-venipuncture variation) for several analytes, and especially for glucose (Figure 1 and Table 3 in Paper II, and Table 3 and 4 in Paper III). Factors that may cause between-venipuncture variation are variation in blood flow, difference in muscle strenght, blood pressure, the position of the arm, the tightness of tourniquet, and the depth and cut of the phlebotomy. Even though the patients in the out-patient clinic often were more difficult to venipuncture than the healthy individual, because of more frequent phlebotomies, fragile veins and hematoma, and difficulties in finding the veins, glucose was the only analyte where the CIs of the between-venipuncture SDs were not overlapping (experiment 1 and 2 in study III) (Table 7). Using butterfly needles as in experiment 2, or venipuncturing healthy individuals (study II), gives the lowest between-venipuncture variation for glucose, which may be caused by a more smooth phlebotomy.

5.2.4 Methodological considerations

The width of the CIs

The estimates of random variation neither increase nor decrease systematically with the number of participants, but the confidence intervals are narrower for larger samples. To decrease the preanalytical SDs, the practice has to improve or change. To

decrease the uncertainty (the width of the CIs) of the preanalytical SDs, the number of participants has to increase. The acceptable magnitude of the preanalytical SD and width of the 95% CI, depend on the analyte and clinical use. Glucose needs accurate estimates because the parameter is used in the diagnosis of diabetes. Reducing the relative width of the CI for preanalytical SD for glucose from 29% to 23-25% when increasing the number of participants from 20 to 30 in study III, does not seem very decisive for the calculation of the total uncertainty (Table 8).

When the preanalytical SDs were estimated separately for each of the paired treatments in experiment 2, for e.g. the green needles (Table 2, Paper III), the test results from both optimally and suboptimally mixed tubes were included. If e.g. the preanalytical SDs for the tubes collected by the green needles should be estimated based on only the results from the optimally mixed tubes, would have required dividing the dataset in experiment 2 in four parts. Because of the risk of losing power, resulting in unstable or wider CIs, we chose to divide the dataset in only two parts. On the other hand, there may be fewer assumptions that can be wrong in more isolated situations. The possibility for separate modeling in different clusters is higher in a large than in a small data set. When the preanalytical SDs were estimated separately for the optimally treated and the suboptimally treated group (Table 2, Paper III), the number of unstable 95% CIs were higher, than when using the whole data set (Table 3 and 4, Paper III).

The reason why the widths of the CIs in most cases are broader for the between-venipuncture SDs (Figure 6) may be that the estimates are based on two venipunctures from each participant, while the preanalytical SDs are based on test results from 4 tubes from each arm, and the measurement repeatability SDs are calculated based on results from duplicates of all tubes. The width of the between-venipuncture CIs would probably be somewhat reduced if e.g. four venipunctures were performed on the same individual.

Skewness

It is important to know the distribution of the test results in order to obtain correct estimates. Excluding the high results in the tail (sensitivity analysis), gives estimates based on less skewed test results. Comparing these estimates with the estimates from the main analysis, which included all the data (as shown in paper III), demonstrate that linear mixed-effects models are pretty robust against skewed data. The estimates from the sensitivity analysis are mostly comparable. The high test results should not be removed from the main analysis, because they may be part of the variables' natural variation.

The model estimates in study II and III will be applicable to the concentration area where the majority of the test results in the studies were. Thus, in concentration areas where there are few data, as in the right tail of the distribution (study III), the estimates will be less applicable. The estimates will not be applicable for concentration areas outside of the area that was the basis for the model.

5.3 Treatment of biases

If a specified treatment gives a well-defined bias and the probability for using the alternative treatment is known, a correction for bias is possible. Though, making corrections is demanding, and requires a high degree of security of the estimated bias, including narrow CIs estimated from a large data set. It will require that the routine practice is similar to the research practice, where the bias was estimated, and standard procedures must be followed. Preanalytical handling will often create biases that are unverifiable and unpredictable, and therefore not suitable for corrections.

If EQA reveals an analytical bias which may influence the trueness, common practice is, if possible, to eliminate the bias, in order to sustain the stability of the accuracy of the method, and the reliability of the reference intervals. It has been discussed whether non-significant analytical bias should be included in the uncertainty budget, because when the bias is non-significant, it is assumed that a procedure is unbiased (74). As a

result it was argued that a non-significant bias should be included in the uncertainty budget, when the uncertainty represents at least 30% of the overall uncertainty (74). As shown in the uncertainty budgets, the biases may also neutralize each other (Table 2, Paper I).

We have presented two cases for combining bias and uncertainty, 1) the MSE budget and 2) an example of recalculation of the bias estimated between ways of transportations, to random variation, by connecting the probability for transporting the samples by the pneumatic tube system. These recalculations are planned to be further elaborated.

In parallel with the quality specifications for analytical performance, there should also be quality specifications for preanalytical performance. The demands on analytical performance derived from biological variation were stated to satisfy general medical needs (38). The biases for all analytes (study I, II, and III), except for glucose, were within the quality specifications. This indicates that, except for glucose, the existing practice at the laboratory using RST tubes, butterfly needles, pneumatic tube transport, and suboptimal mixing, is acceptable when evaluated against these quality specifications.

When the analytical performance fulfils the analytical demands based on biological variation (5), we may experience that when preanalytical uncertainty is included, the demands are not fulfilled. There are guidelines for optimal preanalytical handling (6;13), and two models for estimating preanalytical uncertainty has been established in this work. The quality specifications should be revised to cover the total uncertainty, including both analytical and preanalytical uncertainty.

Critical difference or the reference change value (RCV) is calculated in order to examine whether a specific change in serial test results from a patient, is caused by a change in the medical condition, or may be caused by the within-subject biological variability and the analytical imprecision of the method (5). The formula for the

calculation of RCV involves the analytical variation and the within-subject variation, while the preanalytical variation is considered negligible (5). Results from our studies may indicate that the RCV application should be revised by including the preanalytical variation.

5.4 Traceability, trueness, accuracy

EQA adds substantial value to the practice of laboratory medicine, and global standardization and harmonization of analytes are needed to support clinical practice (44). In EQA, the analytical accuracy is evaluated, not the preanalytical quality. Questionnaires are used in EQA to register the participants' preanalytical practices or solutions on preanalytical problems (16). A working group on the preanalytical phase intends to identify some of the most critical elements in this phase, and to make recommendations on how to reduce the impact of the preanalytical phase (16).

If we should transmit the concept of traceability from the analytical to the preanalytical field, evidence based guidelines such as the CLSI guidelines on recommended phlebotomy and preanalytical sample handling (6;13), can function as a reference method in preanalytical handling, similar to analytical calibrators that are traceable to a reference in the analytical field. Thus, preanalytical traceability would require estimating the preanalytical uncertainty for samples handled optimally, based on recommended guidelines, by using a valid statistical method. This will represent the preanalytical accuracy, the best practice with the minimal preanalytical uncertainty. Similar to analytical variation, the preanalytical uncertainty has to be monitored and verified.

6. MAIN CONCLUSIONS

Two models for estimating preanalytical uncertainty have been established. By the "bottom-up" approach, the uncertainty for each uncertainty source has been estimated, and the individual expected biases and variances combined in an uncertainty budget. In the "top-down" approach, by using linear mixed-effects models, the total preanalytical variation within clustering levels of the data was estimated, together with the biases between different treatments. Both models are practical, and have limited and acceptable methodological assumptions. The model estimates are valid with acceptable size of most confidence intervals.

Glucose was the only analyte falling outside the quality specifications for analytical bias, when the significant biases between different preanalytical treatments in Paper II and III, and the combined biases from the uncertainty budget in Paper I were compared with quality specifications. For glucose, the combined bias from the uncertainty budget (Paper I), and the mean difference between SST vs RST tubes (Paper II) was falling outside the quality specifications, with prolonged clotting and storage time as the greatest contributors to the uncertainty budget. This may have medical consequences. Choice of tube is especially important for glucose, and a standardized procedure is needed in order to reduce the preanalytical uncertainty for glucose.

In Paper II, the preanalytical SDs (excluding the between-venipuncture SD) for LD and potassium for optimally treated samples, were significantly higher than the measurement repeatability SDs, but for glucose, the between-venipuncture SD was the dominant source of variation. The total uncertainty more than doubled, when the preanalytical SDs and the between-venipuncture SDs were added to the measurement repeatability for albumin, ALP, CK, glucose, LD, and potassium (Paper II).

In Paper III, the preanalytical SD for glucose for samples transported in the pneumatic tube system, was significantly higher than for samples manually delivered. Use of different needles and mixing methods gave no important effects.

For most analytes, the preanalytical SDs estimated in study II and III, were about the same, indicating that the preanalytical variations are less influenced by different preanalytical handling, and by participants included in the studies. Consequently, the minimal preanalytical SDs for the analytes, based on optimal phlebotomy and handling of the blood tubes collected from healthy individuals, may be considered as a reasonable estimate of the preanalytical variation of blood samples, also for blood samples collected in an outpatient clinic, and suboptimally treated as shown.

Overall, a standardized preanalytical handling procedure based on guidelines should be settled for each component. The estimation of the minimal preanalytical uncertainty based on optimal handling of the blood samples is important, especially for components where accurate estimates of the test results are needed in clinical practice. Establishing an uncertainty budget is especially important when using treatments alternative to optimal treatment. Estimation of preanalytical uncertainty may improve diagnostic quality and patient treatment.

7. FUTURE PERSPECTIVES

Further work on recalculation of biases to random variation is planned.

We would like to use linear mixed-effects models further in estimating the minimal preanalytical uncertainty for even more analytes, and also introduce new uncertainty sources such as e.g. posture during phlebotomy, and resting in a sitting position for 15 min before phlebotomy. Preanalytical uncertainty may be estimated using samples from clinically relevant situations as e.g. heart patients, when testing on heart markers.

The cause of the between-venipuncture variation should be elucidated by further research.

Based on our models, it will be considered at our laboratory, to start the process writing a guideline, on how to estimate preanalytical uncertainty. Estimation of preanalytical uncertainty may be a part of the general method validation to improve the quality of the whole analysing process.

Internationally, preanalytical traceability may be defined for each analyte, a stronger emphasis on standardization of preanalytical treatment may be needed, and quality specifications should be made for preanalytical uncertainty.

References

1. Joint Committee for Guides in Metrology. Evaluation of measurement data - guide to the expression of uncertainty in measurement. 1st ed., 2008; corrected version, 2010. JCGM 100: 2008 (GUM 1995 with minor corrections). http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (Accessed February 2013): 2010:120pp.

2. Clinical and Laboratory Standards Institute. Expression of measurement uncertainty in laboratory medicine; approved guideline. Document nr C51-A. Wayne (PA): CLSI, 2012:57pp.

3. Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. Clinical Chemistry 1997;43:1348-51.

4. Plebani M. Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med 2006;44:750-9.

5. Fraser CG. Biological variation: From principles to practice. United States of America. AACCPress, 2001:140-141pp.

6. Clinical and Laboratory Standards Institute. Procedures for the collection of diagnostic blood specimens by venipuncture; approved standard - sixth edition. Document nr H3-A6. Wayne (PA): CLSI, 2007:56pp.

7. Joint Committee for Guides in Metrology. International vocabulary of metrology - basic and general concepts and associated terms (VIM), 3rd edition, JCGM 200:2012. http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf (Accessed February 2013): 2012:91pp.

8. Clinical and Laboratory Standards Institute. Laboratory instrument implementation, verification, and maintenance; approved guideline - fourth-edition. Document nr GP31-A. Wayne (PA): CLSI, 2009:71pp.

9. Klee GG, Schryver PG, Kisabeth RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. Scand J Clin Lab Invest 1999;59:509-12.

10. Lippi G, Chance JJ, Church S, Dazzi P, Fontana R, Giavarina D et al. Preanalytical quality improvement: from dream to reality. Clin Chem Lab Med 2011;49:1113-26.

11. Guder WG, Narayanan S, Wisser H, Zawta B. Diagnostic samples: from the patient to the laboratory: the impact of preanalytical variables on the quality of

laboratory results, 4 ed. Weinheim: Wiley-VCH Verlag GmbH & Co. KgaA, 2009:113pp.

12. Oosterhuis WP, Bruns DE, Watine J, Sandberg S, Horvath AR. Evidence-based guidelines in laboratory medicine: Principles and methods. Clin Chem 2004;50:806-18.

13. Clinical and Laboratory Standards Institute. Procedures for the handling and processing of blood specimens for common laboratory tests; approved guideline - fourth-edition. Document nr H18-A4. Wayne (PA): CLSI, 2010:57pp.

14. Clinical and Laboratory Standards Institute. Tubes and additives for venous and capillary blood specimen collection; approved standard - sixth-edition. Document nr H01-A6. Wayne (PA): CLSI, 2010:17pp.

15. Clinical and Laboratory Standards Institute. Validation and verification of tubes for venous and capillary blood specimen collection; approved guideline. Document nr GP34-A. Wayne (PA): CLSI, 2010:45pp.

16. Lippi G, Becan-McBride K, Behulova D, Bowen RA, Church S, Delanghe J et al. Preanalytical quality improvement: in quality we trust. Clin Chem Lab Med 2013;51:229-41.

17. Ramsey MH, Thompson M. Uncertainty from sampling, in the context of fitness for purpose. Accred Qual Assur 2007;12:503-13.

18. Eurachem/CITAC. Measurement uncertainty arising from sampling. A guide to methods and approaches. First ed. Eurachem, 2007:111pp.

19. ISO Standards. Medical laboratories - Particular requirements for quality and competence (ISO 15189:2007). Genève: ISO Standards, 2007:39pp.

20. Lippi G, Salvagno GL, Montagnana M, Brocco G, Guidi GC. Influence of the needle bore size used for collecting venous blood samples on routine clinical chemistry testing. Clin Chem Lab Med 2006;44:1009-14.

21. Parenmark A, Landberg E. To mix or not to mix venous blood samples collected in vacuum tubes? Clin Chem Lab Med 2012;49:2061-3.

22. Streichert T, Otto B, Schnabel C, Nordholt G, Haddad M, Maric M et al. Determination of hemolysis thresholds by the use of data loggers in pneumatic tube systems. Clin Chem 2011;57:1390-7.

23. Fuentes-Arderiu X, Acebes-Frieyro G, Gavaso-Navarro L, Castiñeiras-Lacambra MJ. Pre-metrological (pre-analytical) variation of some biochemical quantities. Clin Chem Lab Med 1999;37:987-9.

24. García-Panyella M, Padró-Miquel A, Dot-Bach D, Fuentes-Arderiu X. Pre-analytical variation of some haematological quantities. Clin Chem Lab Med 2008;46:1168-70.

25. Padró-Miquel A, García-Panyella M, García-Giménez MJ, Fuentes-Arderiu X. Pre-metrological variation of immunological quantities. Scan J Clin Lab Invest 2008;68:431-2.

26. Kallner A, Waldenström J. Does the uncertainty of commonly performed glucose measurements allow identification of individuals at high risk for diabetes? Clin Chem Lab Med 1999;37:907-12.

27. Kouri T, Siloaho M, Pohjavaara S, Koskinen P, Malminiemi O, Pohja-Nylander P, Puukka R. Pre-analytical factors and measurement uncertainty. Scan J Clin Lab Invest 2005;65:463-75.

28. Linko S, Örnemark U, Kessel R, Taylor PDP. Evaluation of uncertainty of measurement in routine clinical chemistry - Applications to determination of the substance concentration of calcium and glucose in serum. Clin Chem Lab Med 2002;40:391-8.

29. Kristiansen J. The Guide to Expression of Uncertainty in Measurement approach for estimating uncertainty: An appraisal. Clin Chem 2003;49:1822-9.

30. Kirkup L, Frenkel RB. An introduction to uncertainty in measurement using the GUM (guide to the expression of uncertainty in measurement). Cambridge: Cambridge University Press, 2006:233pp.

31. Eurachem/CITAC. Quantifying uncertainty in analytical measurement. 3rd. ed. http://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf (Accessed November 2012). London: Eurachem, 2012:141pp.

32. National Institute of Standards and Technology. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Gaithersburg: NIST, 1994:20pp.

33. BIPM (Bureau International des Poids et Mesures) - Bibliography on uncertainty http://www.bipm.org/en/publications/guides/wg1_bibliography.html (Accessed January 2013).

34. van der Veen AMH, Cox MG. Error analysis in the evaluation of measurement uncertainty. Metrologia 2003;40:42-50.

35. Kristiansen J. Description of a generally applicable model for the evaluation of uncertainty of measurement in clinical chemistry. Clin Chem Lab Med 2001;39:920-31.

36. Krouwer JS. Critique of the guide to the expression of uncertainty in measurement method of estimating and reporting uncertainty in diagnostic assays. Clin Chem 2003;49:1818-21.

37. Hyltoft Petersen P, Fraser CG, Kallner A, Kenny D. Strategies to set global analytical quality specifications in laboratory medicine. Scand J Clin Lab Invest 1999;59.

38. Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491-500.

39. Biological variation and desirable specifications for QC http://www.westgard.com/guest17.htm (Accessed February 2013).

40. Phillips SD, Eberhardt KR, Parry B. Guidelines for expressing the uncertainty of measurement results containing uncorrected bias. J Res Natl Inst Stand Technol 1997;102:577-85.

41. Vesper HW, Thienpont LM. Traceability in Laboratory Medicine. Clin Chem 2009;55:1067-75.

42. Little RR, Rohlfing CL, Wiedmeyer HM, Myers GL, Sacks DB, Goldstein DE. The National Glycohemoglobin Standardization Program: A five-year progress report. Clin Chem 2001;47:1985-92.

43. CLIA criteria of analytical performance. Federal Register February 28, 1992;57(40):7002-186. http://www.westgard.com/clia.htm#chem (Accessed February 2013).

44. Miller WG, Jones GRD, Horowitz GL, Weykamp C. Proficiency Testing/External Quality Assessment: Current Challenges and Future Directions. Clin Chem 2011;57:1670-80.

45. Ricós C, Ramón F, Salas A, Buño A, Calafell R, Morancho J et al. Minimum analytical quality specifications of inter-laboratory comparisons: agreement among Spanish EQAP organizers. Clin Chem Lab Med 2012;50:455-61.

46. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag, 2000:608pp.

47. Linnet K, Boyd JC. Selection and analytical evaluation of methods - with statistical techniques. In: Burtis CA, Ashwood ER, Bruns DE, eds. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics. Philadelphia: Elsevier Saunders, 2011:7-47.

48. Davis CS. Statistical methods for the analysis of repeated measurement. New York: Springer-Verlag, 2002:417pp.

49. Lippi G, Salvagno GL, Brocco G, Guidi GC. Preanalytical variability in laboratory testing: influence of the blood drawing technique. Clin Chem Lab Med 2005;43:319-25.

50. Bowen RAR, Hortin GL, Csako G, Otañez OH, Remaley AT. Impact of blood collection devices on clinical chemistry assays. Clin Biochem 2010;43:4-25.

51. Chu SY, Macleod J. Effect of 3-Day Clot Contact on Results of Common Biochemical Tests with Serum. Clin Chem 1986;32:2100.

52. Laessig RH, Indriksons AA, Hassemer DJ, Paskey TA, Schwartz TH. Changes in serum chemical values as a result of prolonged contact with clot. Am J Clin Pathol 1976;66:598-604.

53. Martin BJ, Mcgregor CW. Measurement of Serum Magnesium - Effect of Delay in Separation from Erythrocytes. Clin Chem 1986;32:564.

54. Sacks DB. Carbohydrates. In: Burtis CA, Ashwood ER, Bruns DE, eds. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics. Philadelphia: Elsevier Saunders, 2011:709-30.

55. Cui M, Jing RR, Wang HM. Changes of serum lactate dehydrogenase and potassium levels produced by a pneumatic tube system. Labmedicine 2009;40:728-31.

56. Felder RA. Preanalytical errors introduced by sample-transportation systems: a means to assess them. Clin Chem 2011;57:1349-50.

57. Koessler J, Kobsar AL, Brunner K, Stolz H, Dossler B, Walter U, Steigerwald U. The preanalytical influence of two different mechanical transport systems on laboratory analysis. Clin Chem Lab Med 2011;49:1379-82.

58. Sodi R, Darn SM, Stott A. Pneumatic tube system induced haemolysis: assessing sample type susceptibility to haemolysis. Ann Clin Biochem 2004;41:237-40.

59. Steige H, Jones JD. Evaluation of pneumatic-tube system for delivery of blood specimens. Clin Chem 1971;17:1160-64.

60. Wallin O, Söderberg J, Grankvist K, Jonsson PA, Hultdin J. Preanalytical effects of pneumatic tube transport on routine haematology, coagulation parameters, platelet function and global coagulation. Clin Chem Lab Med 2008;46:1443-9.

61. Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. Clin Chem Lab Med 2008;46:764-72.

62. Caraway WT. Chemical and diagnostic specificity of laboratory tests - effect of hemolysis, lipemia, anticoagulants, medications, contaminants, and other variables. Am J Clin Pathol 1962;37:445-64.

63. Söderberg J, Jonsson PA, Wallin O, Grankvist K, Hultdin J. Haemolysis index - an estimate of preanalytical quality in primary health care. Clin Chem Lab Med 2009;47:940-4.

64. Banks RE, Stanley AJ, Cairns DA, Barrett JH, Clarke P, Thompson D, Selby PJ. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem 2005;51:1637-49.

65. Crawley MJ. Statistics An introduction using R. Chicester: John Wiley & Sons Ltd., 2010:327pp.

66. Bland JM, Altman GG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-60.

67. Bain LJ, Engelhardt M. Introduction to probability and mathematical statistics, 2nd ed. Belmont, Calif: Duxbury Press, 1991:656pp.

68. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York: Springer-Verlag, 2001:528pp.

69. Heydorn K, Madsen BS. Verification of uncertainty budgets. Accred Qual Assur 2005;10:403-8.

70. Zhang DBJ, Elswick RK, Miller WG, Bailey JL. Effect of serum-clot contact time on clinical chemistry laboratory results. Clin Chem 1998;44:1325-33.

71. Clinical and Laboratory Standards Institute. Verification of comparability of patient results within one health care system; approved guideline. Document nr C54-A. Wayne (PA): CLSI, 2008:49pp.

72. Risteli J, Winter WE, Kleerekoper M, Risteli L. Bone and mineral metabolism. In: Burtis CA, Ashwood ER, Bruns DE, eds. Tietz textbook of clinical

chemistry and molecular diagnostics. 5th Ed. Philadelphia: Elsevier Saunders, 2012:1733-801.

73. Carraro P, Servidio G, Plebani M. Hemolyzed specimens: A reason for rejection or a clinical challenge? Clin Chem 2000;46:306-7.

74. Maroto A, Boque R, Riu J, Rius FX. Should non-significant bias be included in the uncertainty budget? Accred Qual Assur 2002;7:90-4.