External Quality Assurance of Point-of-Care testing

Participant performance and commutability of the control material

Tone Bukve

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

External Quality Assurance of Point-of-Care testing

Participant performance and commutability of the control material

Tone Bukve

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 03.09.2021

Year: 2021

Title: External Quality Assurance of Point-of-Care testing

Name: Tone Bukve

Print: Skipnes Kommunikasjon / University of Bergen

Department of Global Public health and Primary Care

Faculty of Medicine and Dentistry

University of Bergen

Norwegian quality improvement of primary care laboratories

Bergen

In memory of my father, Odd Bukve

Scientific environment

This thesis is a part of the PhD program and the research group *"Kunnskapsbasert bruk og nytte av laboratorienanlyser"*. The research group is located at Noklus, Haraldsplass Deaconess Hospital, Bergen. The group members are professor dr. med. Sverre Sandberg (director), PhD Anne Vegard Stavelin, professor dr. med. Svein Skeie, Phd Una Sølvik, PhD Gunn BB Kristensen, PhD Ann-Helen Kristoffersen, PhD Åsne Karine Aarsand, PhD Mette Christophersen Tollånes, Pernille Kjeilen Fauskanger (statistican), Gro Gidske, PhD Pia Borgen Gjerde, professor Pål Arne Jenum, and md. Svein Ivar Fylkesnes. The research group at Noklus is located at Haraldsplass Deaconess Hospital and cooperates with the Department of Global Public health and Primary Care. I was invited into the group at Noklus through participation in the research educational program at the Faculty of Medicine. Professor Sandberg has been my main supervisor and I have been co-supervised by Anne Stavelin. A substantial part of the thesis has been performed in collaboration with the abovementioned colleagues at Noklus, in addition to Berit Riksheim, Nina Christensen, and Wenche Vie, all employees in the EQA department of Noklus.

Acknowledgements

Although completing this PhD has been a challenging task, it has been a journey that I have truly enjoyed. I owe many thanks to the wonderful people who have helped me along the way.

First of all, I would like to thank my main supervisor, Sverre Sandberg, who introduced me to the world of laboratory medicine and research when I was still a student at the Faculty of Medicine. Our first meeting was in your office at Noklus, the only time you had available that day, at nine o'clock in the evening. As you did then, no matter how busy your schedule, you have always taken the time to meet with me to provide feedback, support, and advice. I have greatly benefitted from your extensive knowledge and experience, and I am very grateful for all the work you have done. Finally, I must thank you for your kindness and understanding, and for supporting me through the years so that I was able to finish my PhD.

Anne Stavelin, my highly valued co-supervisor, thank you for sharing your knowledge, your constructive critiques, and your excellent advice. Your assistance throughout the project has been indispensable. I am very grateful that you were always there to lend me a helping hand and that you kept your office door open so that I could stop by whenever needed.

I would also like to extend my gratitude to my coauthors, it has been a true privilege working with all of you. A special thanks to Berit Oddny Riksheim for your support and to Wenche Vie for so generously including me in your work on commutability. Thanks also to all my colleagues at Noklus for making it such an inspiring and inclusive workplace.

Finally, I would like to thank my family. My parents-in-law, Kari and Rolf, thank you for all your support and care. Thank you to my sister Trude, my brother Tord, and their significant others, Gunnar and Guro, for always cheering me on. To my mother Gunn Wenche, I am so grateful that you always believe in me, as well as for your self-sacrifice and unconditional love. You have always helped me see the bigger picture

when I find myself distracted by the details. A special thank you to my dear father, Odd, as I find myself preparing for the final step toward my degree without him by my side. He was my motivation and my rock. His broad knowledge, strong sense of justice, stubbornness, loyalty, kindness, and love empower me every day. Just like my sister, I dedicate the thesis to him. I end by thanking my own “little gang”, Ole and our kids Ane, June, Kyrre, and Sondre. Life is good spending it with people who make you happy - you are my life.

Abstract

The term point-of-care (POC) testing refers to laboratory tests performed near the patient, as opposed to a clinical laboratory setting. The aim of POC testing is to generate fast test results that facilitate accurate diagnosis and the initiation of proper treatment. POC testing is intended to improve patient outcomes and make laboratory testing available in settings where it has previously been unavailable. The measurement procedures (MPs) used for POC testing can be either simple test strips or complex analytical systems. In general, the MPs are more or less fully automated and often require few steps after sample collection before the test results are obtained. Since POC testing is usually performed outside the central laboratories, it is likely to be operated by persons who have not been trained as laboratory scientists, such as doctors, nurses, other healthcare professionals, or even the patients themselves.

POC testing is the fastest-growing part of laboratory testing and makes up a notable share of the in vitro diagnostic medical device market. While POC testing eliminates some of the more problematic steps in the testing process, such as transportation of samples and distribution of test results, new challenges are introduced. Specifically, concerns regarding the test quality of POC testing have emerged, however, they can be addressed through the quality management of POC testing. In Norway, users of POC testing participate in the Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), which provides a quality management system, including external quality assurance (EQA), covering the total testing process.

Being a fundamental part of the EQA scheme, the EQA material must have characteristics that allow for evaluating the different MPs and their accuracy, precision, and calibration traceability, as well as provide a collective assessment of the laboratories, evaluating their reproducibility, standardization, and overall harmonization. The EQA material must possess a range of qualities to fulfill these requirements, the most important being commutability.

The aim of this thesis was to evaluate the effect of participating in an EQA organization and to evaluate the MPs and EQA material used to improve the participants' analytical quality.

In papers I and II, the aim was to evaluate whether the analytical performance of POC testing was influenced by participation in a quality improvement system over time. The factors related to good participant performance were also addressed. The analytes included in the two studies were urine albumin (u-albumin), C-reactive protein (CRP), glucose, and hemoglobin (Hb). Results from EQA schemes at Noklus were assessed, 15 and 19 EQA schemes from 1998 to 2012 and 2006 to 2015, respectively. In paper II, logistic regression was performed to enable assessment of which characteristics predicted good participant performance. The average numbers of participants in the EQA schemes were 1159, 2134, 2357, and 2271 for u-albumin, CRP, glucose, and Hb, respectively. The results from the two studies showed a gradual increase in the percentage of participants with good performance. Furthermore, paper II showed that performing weekly internal quality control, performing the laboratory analysis ten or more times per week, using recommended MPs, and having operators that were qualified in laboratory medicine were all predictors of good participant performance in the EQAs.

In paper III, the aim was to assess whether Noklus' in-house whole blood EQA material for CRP, glucose, and Hb was commutable for the most frequently used POC MPs in Norway and assess the possibility of using a common target value for each of the three analytes. The study was performed following the guideline of the Clinical and Laboratory Standards Institute, EP14-A3 evaluation of commutability of processed samples, 3rd edition. The EQA material was pooled stabilized EDTA venous whole blood containing different concentrations of CRP, glucose, and Hb. The EQA material and native routine patient samples were analyzed using 17 POC and 3 hospital MPs. Deming regression with 95% prediction intervals for each MP comparison was used for assessing commutability. The evaluation showed that the EQA material was commutable for all the CRP and Hb POC MPs and all glucose POC MPs at the lowest

concentration (126.0 mg/dL (7.0 mmol/L)) and for the three glucose POC MPs at all concentrations.

In conclusion, the studies in this thesis have revealed that systematic participation in a quality management system for POC testing in primary healthcare can improve the analytical quality of u-albumin, CRP, Hb, and glucose measurements. The whole blood EQA material used was commutable for all CRP and Hb MPs, and for approximately half of the glucose MPs used in primary healthcare in Norway. Both participant and MP evaluations can now be performed for these participants and MPs.

List of Publications

Paper I

Bukve T, Røraas T, Riksheim BO, Christensen NG, Sandberg S.

Pont-of-care urine albumin in general practice offices: effect of participation in an external quality assurance scheme.

Clin Chem Lab Med 2015;53:45-51. https://doi.org/10.1515/cclm-2014-0483.

Paper II

Bukve T, Stavelin A, Sandberg S.

Effect of Participating in a Quality Improvement System over Time for Point-of-care C-Reactive Protein, Glucose, and hemoglobin Testing.

Clin Chem 2016;62:1474-1481. https://doi.org/10.1373/clinchem.2016.259093.

Paper III

Bukve T, Sandberg S, Vie WS, Sølvik U, Christensen NG, Stavelin A.

Commutability of a whole-blood external quality assurance material for point-of-care C- reactive protein, glucose, and hemoglobin testing.

Clin Chem 2019;65:791-797. https://doi.org/10.1373/clinchem.2018.300202.

List of abbreviations and definitions

Abbreviations

APS: Analytical performance specifications

CI: Confidence interval

CLSI: Clinical and Laboratory Standards Institute

CRM: Certified reference material

CRP: C-reactive protein

CV: Coefficient of variation

EDTA: Ethylenediamine tetraacetic acid

EFLM: European Federation of Clinical Chemistry and laboratory Medicine

EQA: External quality assurance

EU: European Union

GP: General practitioner

Hb: Hemoglobin

HiCN: Cyanmethemoglobin

ICSH: International Council for Standardization in Hematology

ID GC-MS: Isotope dilution mass spectrometry

IFCC: International Federation of Clinical Chemistry

IQC: Internal quality control

ISO: International Organization for Standardization

IVD: In vitro diagnostic

JCTLM: The Joint Committee for Traceability in Laboraory

MP: Measurement procedure

Noklus: The Norwegian Organization for Quality Improvement of Laboratory Examinations

OR: Odds ratio

PhD: Philosophiae doctor

PI: Prediction interval

POC: Point-of-care

TAT: Turnaround time

TE: Total error

TE_a: Allowable total error

SD: Standard deviation

SI: Systeme Internationale

SKUP: The Scandinavian evaluation of laboratory equipment for point of care testing

U-albumin: Urine albumin

Definitions

Accuracy: How close a single measurement value is to a real value. Accuracy is related to both the trueness and precision of a measurement value.

Measurement error: Measured value minus a reference value.

Random measurement error: How the measurement error between replicates varies unpredictably. The random error can qualitatively be expressed as precision.

Precision: The closeness of agreement between replicate measurement values obtained under specific conditions. Precision is quantitively expressed as imprecision such as standard deviation or in percentage as the coefficient of variation under specific conditions.

The specific conditions associated with precision/imprecision are repeatability, intermediate precision, and reproducibility, listed with respect to an increasing degree of variability. The specific conditions may also be divided into imprecision within one laboratory (compromising the repeatability and intermediate precision) and imprecision between laboratories (reproducibility).

Repeatability: Replicate measurements on the same or similar object under the same set of conditions including the same measurement procedure, same operators, same conditions, and same location. The replicates are measured over a short period of time, usually the same day, (within-series variation).

Intermediate precision: Replicate measurements on the same or similar object under the same set of conditions including the same measurement procedure, and same location, but may also involve changes in condition. The replicates are measured over an extended time-period, (within-laboratory variation).

Reproducibility: Replicate measurement on the same or similar object under a different set of conditions including different measurement procedures, different operators, and different locations, (between-laboratory variation).

Trueness: Closeness of agreement between the average of a number of replicates measurement values and a reference value. Trueness is inversely related to *bias*.

Bias/systematic measurement error: A measurement error which in replicate measurements is constant or varies in a predictable manner.

Confidence interval: An interval associated with a confidence level representing the degree of uncertainty of the calculated value.

Prediction interval: An interval where a future value is most likely to fall.

Standardization: That measurement results are traceable, via calibration, to a higher-order reference material or a reference measurement procedure.

Harmonization: Measurement results being equivalent without traceability to a reference measurement procedure and not emphasizing the process to achieve equivalent results.

Reference measurement procedure: A measurement procedure yielding measurement results fit for intended use when assessing measurement trueness obtained from other measurement procedures.

Reference material: A material which is adequately homogenous and stable and established as fit for intended use in a measurement.

CRM: A reference material accompanied by a certificate from an authoritative body providing specified property values, its associated uncertainty, and a statement of its traceability.

Calibration hierarchy: A sequence of calibrations from a reference to a final measuring system.

Primary measurement procedure: A measurement procedure acquiring the highest metrological characteristics, i.e., affecting both the estimate of the measurand and any uncertainty of this estimate (reported in SI units), additionally, the measurement procedure’s operation is described in detail and fully understood.

Secondary reference measurement procedure: A measurement procedure with high selectivity for the analyte and thereby being minimally influenced by any matrix components in the sample, in addition, the procedure should have high precision.

Primary reference material: Is an ultra-pure material which contains the relevant analyte in a certified amount, any impurities are accounted for, and the material is stable and traceable.

Secondary reference material: A reference material which often has a matrix close to that of clinical samples, or the material can be composed of a panel of clinical samples.

Commutability: The processed sample tested has the same mathematical relationship between measurement procedures as is observed for representative authentic clinical samples.

Matrix-related bias: Bias not normally seen in in authentic clinical samples but often in processed samples.

Sample matrix: All other substances in the material but the measurand itself.

The definitions are modified from VIM, Bolann, and Miller et al. (1-3).

List of tables

List of figures

Contents

1. Introduction

In clinical chemistry, analyses of body fluids are performed to yield relevant, accurate, and precise information about the health status of the patient. The cornerstone of all laboratory medicine is that the in vivo processes in the human body can be understood by analyzing body fluids in vitro (4). The modern laboratory is a physical place fully equipped with complex measurement procedures (MPs) and staffed with professionally trained operators who analyze hundreds or even thousands of specimens daily (4). However, the history of laboratory testing started as isolated testing at the patient site with small and simple MPs, which were easy to use (4). Today near patient testing, or point-of-care (POC) testing, is again increasing as health services are becoming more patient centered, and it is now the fastest growing part of the laboratory testing market (5).

External quality assurance (EQA) was launched over 70 years ago to address the observation that laboratory results for aliquots of the same sample showed different results when analyzed at different laboratories (6, 7). The results obtained from the EQA schemes were used to attain more uniform laboratory results among the laboratories. Today, EQA schemes are a crucial component of laboratories' quality management systems, assessing both the analytical and the pre-and postanalytical performance of the laboratory. The EQA surveys are managed by distributing EQA samples with unknown concentrations to the participants. One of the most important properties of the EQA material is commutability, which affects both the design and the interpretation of the EQA results.

The present thesis evaluates the commutability of EQA materials and the effect of participating in a quality improvement system for POC testing over time.

1.1 Point-of-care testing

POC testing is a term used to describe laboratory tests carried out near the patient, rather than in the clinical laboratory setting. The aim of performing POC testing is to make laboratory test results available quickly and thereby expedite diagnosis and initiate proper treatment, with the intention of improving patient outcomes (8). The key features of POC testing include no sample preparation, no pipetting, reagents that are ready to use, no special medical technical background is needed to perform the tests, the instant availability of results, and the initiation of diagnosis and treatment without delay. POC testing is used in a variety of settings both within hospitals, such as in the intensive care unit, in the diabetic care ward, and in the operating and delivery rooms, as well as outside the hospital setting, such as the offices of general practitioners (GPs) and in home care for self-monitoring performed by the patients. There are currently POC tests for a broad range of clinical parameters, and new tests, such as the viral tests for influenza and Covid-19, are continuously being developed (9, 10). As no uniform term is agreed upon for this type of testing, POC testing is also termed “near patient testing”, “remote rapid testing”, “decentralized testing” or “bedside testing” (11).

The MPs used for POC testing span from simple test strips to complex analytical systems. The MPs are, in general, more or less fully automated, and they tend to require very few steps from the point at which the sample is collected to the generation of the final test result. Different designs for POC MPs have entered the market, benchtop analyzers, which are complex analytical systems often used at GP offices, emergency departments, or in nursing homes, and hand-held MPs, which are mainly used at the bedside in hospitals or in the home environment. The MPs can provide either qualitative or quantitative test results. When using qualitative tests, the results are presented on an ordinal scale and on an interval or ratio scale when using quantitative tests. Hand-held POC testing is often performed using test strips, and the results can either be read directly or by using a simple MP. The analytical reaction for such MPs typically takes place inside the test strips, while the POC MP only generates a readable result. Benchtop analyzers are generally smaller versions of the analytical systems used

in central laboratories, in addition to being more compact, they are also faster and more user friendly (12).

Since POC testing is performed outside central laboratories, these tests are more likely to be carried out by persons other than laboratory scientists, such as nurses, doctors, and other healthcare professionals, or sometimes, the patients themselves (13). However, there are concerns that POC testing performed by non-laboratory staff in a hectic clinical setting is especially prone to errors even though the POC MPs are easy to handle and vulnerable steps in the analyzing process, such as calibration and quality control, may have been automated (13).

POC MPs in Europe are under governmental regulation by the European Union (EU) regulatory framework for in vitro diagnostic medical devices (IVDs). The current framework for POC MPs that are CE marked, the IVD Directive 98/79/EC (14), is currently being replaced by the new EU regulation, Regulation 2017/746, which will become effective from 2022. According to the IVD Directive 98/79/EC, all POC MPs are classified as IVDs, and they must fulfill the requirements regarding quality, safety, and performance, whether they are used in clinical laboratories by trained laboratory scientists, on the ward, or in the home environment by laypersons (15). IVDs are classified into four groups, “high-risk devices”, “risk devices”, “self-testing” (except self-testing for blood-glucose), and “others”, which have the lowest hazard potential. The different groups have different requirements stipulated by the Directive, which determine the design, production, and verification needed for the MPs. Most POC MPs are considered part of the “other” group, which does not include obligations from the manufacturer or a notified body to evaluate the MPs’ quality assurance system. The “self-testing” category has quite strict safety requirements, as the analyses are performed by laypersons outside the laboratories. In contrast, the POC MPs in the “other group” do not yet have any comparable safety demands (15). Drafts by international and European non-governmental standards organizations, such as the European Committee for Standardization and the International Organization for Standardization (ISO), can be adopted to technically specify the different requirements presented in the Directive. There are several standards that apply to POC MPs.

However, they are laid out for POC use in hospitals, by ambulance staff, or in the home environment by patients undertaking self-testing, whereas none of the standards address POC testing performed by non-laboratory staff in the community setting (16). Additionally, standards and guidelines can be acquired from other accredited organizations, such as the Clinical and Laboratory Standards Institute (CLSI), which drafts procedures intended for application globally using a recognized voluntary consensus process (17).

The introduction of POC has the potential to innovate prevention, diagnosis, and monitoring of patients by shortening the turnaround time (TAT) of tests, i.e., the time taken from ordering of a test to when the result is received. The shortened TAT is due to eliminating the time required for both sample transport and sample preparation since the tests are performed at the patient site mainly using whole blood. The immediate diagnosis and shortened therapeutic process aim to reduce the time spent in hospitals, lower the cost of treatment, and improve the satisfaction of patients and staff (18). In developing countries, POC testing has the obvious potential to enable more patients to obtain access to laboratory tests in non-urban areas, where premises for running conventional laboratories are scarce. For example, studies performed in such settings have shown that introducing POC testing in the case of malaria testing has reduced the amount of anti-malaria drugs prescribed to persons with negative malaria test results (19). In the hospital setting, POC tests carried out in the emergency department, where a rapid diagnosis and patient triage are essential, should be advantageous. The usefulness depends, nevertheless, on the TAT of the central laboratory, the measurand, and whether the TAT is the time-limiting step in the particular setting (20). In the wards, a laboratory result from the hospital laboratory, available in a couple of hours, might be satisfactory to provide sufficient medical decision making. POC testing in the operating theatre may shorten the time in surgery or improve the success of operations, e.g., measuring the concentration of parathyroid hormone can facilitate parathyroidectomy (21). In hospital settings, the coordination of laboratory tests performed across different departments is especially important to prevent over-testing, which can increase costs. In outpatient settings, such as GP offices, a quick laboratory

result enables test results to be ready during a consultation. Consequently, the results can be discussed with the patient in the consultation, which has the potential to increase compliance, and enable the initiation of the relevant treatment without delay (22). Therefore, the wait for the patients is shortened, and they do not need to return for a follow-up consultation due to delayed laboratory results. There is, however, limited knowledge on whether shortening the TAT in GP offices actually improves clinical outcomes (23). Still, some studies have shown that using POC testing for patients with suspected urinary tract infections lowered antibiotic use, and testing patients with bacterial infections for C-reactive protein (CRP) resulted in earlier treatment (24, 25). Improved adherence to treatment can be an important advantage of POC testing and has been shown to be effective in cases such as self-testing in patients receiving anticoagulants (26). In many cases, choosing to implement POC testing can be as much an organizational choice as a medical one. For instance, in some areas in Norway, the distance to the nearest hospital laboratory is considerable, which makes POC testing useful, as it enables timely analysis of important analytes, such as CRP (27). In the Netherlands, POC was introduced because of financial cutbacks, which led to more care at the GP office level and downsizing of care offered in hospitals (27). The effect of using POC in primary health care can, however, be difficult to assess since many tests are linked to monitoring chronic diseases, and the impact of the testing might not be measurable before a substantial amount of time has passed.

The use of POC testing is constantly growing, and POC MPs make up a significant share of the IVD market (28). As the number of POC MPs increases, so does the risk of reduced analytical quality of some MPs. In hospitals, systematic differences between the central laboratory and POC MPs can be recognized if the central laboratory supervises the POC MPs. In the outpatient setting, differences between POC MPs used in different locations can be identified when participating in EQA schemes.

1.2 Quality of laboratory testing

1.2.1 The laboratory test process

In 1981, Lundberg (29) introduced a new paradigm for the complete laboratory test process. In his approach, the testing process is organized in a loop, which begins with the physician deciding to request a laboratory test and ends with the physician deciding what measures to take after receiving the test result. Lundberg proposed splitting the test process into categories based on the laboratory testing order. The first category, the pre-preanalytical phase, involves the selection and ordering of laboratory tests according to the patient's medical history and the clinical examination. The preanalytical phase comprises the identification and preparation of patients, collection and handling of samples, and possible transportation and storage of test materials. The analytical phase covers the sample analysis, and the postanalytical phase comprises the validation and reporting of the results. The post-postanalytical phase considers the interpretation of test results and clinical decision making.

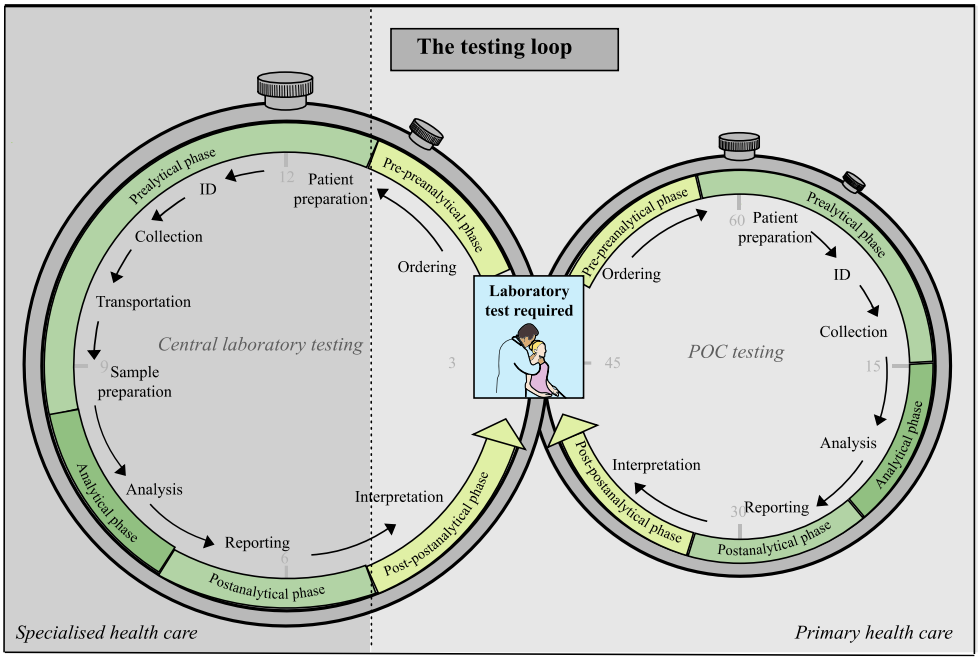


Figure 1. The testing loop. The figure illustrates the clinical sample's lifespan from a laboratory test is required until the test result is received. The clocks represent the time difference between POC and the central laboratory testing. The green colors represent the test phases and the text and arrows the testing process. The figure is modified from Lundberg's figure "Brain to brain loop" (29).

1.2.2 Laboratory errors

Medical errors are preventable, unfavorable events associated with medical care (30). A recent publication estimated that such errors cause the death of about 5000 patients in the US each year (31). Laboratories are responsible for their share of medical errors, however, these errors are generally given less public attention (32). This may be because laboratories make relatively few mistakes overall compared to other medical fields, or it may be that laboratory errors are difficult for non-laboratory personnel to detect (32). Additionally, the laboratory testing process is complex with several steps involved, and the time from ordering a test to the interpretation of the results makes it hard to pinpoint the time and place of the error (32).

Lundberg's categories can be used to help identify and group errors in the laboratory test process. The first and last phases of Lundberg's testing loop are the pre-preanalytical and post-postanalytical phases, both of which are located outside the laboratory. The remaining phases, the preanalytical, the analytical, and the postanalytical, are laboratory phases, and any errors here are therefore categorized as laboratory errors. Nevertheless, in many cases, only the analytical phase is fully controlled by the laboratory since the other phases often involve several different stakeholders (33). Studies have indicated that the pre- and postanalytical errors account for 60% and 20% of the laboratory test errors, respectively (33, 34). However, these errors are mainly diagnostic, including failure to use the indicated test, failure to follow up on the results, or diagnostic delays (35, 36).

1.2.3 Errors in POC testing

In the pre-preanalytical and post-postanalytical phases, errors are associated with ordering the most suitable laboratory tests and interpreting test results, respectively.

In the preanalytical phase, errors can be caused by misidentification of patients or unsuitable patient preparation (for example, in terms of medication, diet, or time of sample collection). Errors can also be related to sample instability as a result of conditions during transport and storage. POC testing, however, eliminates the uncertainty due to stability since tests are performed at the patient site immediately after collection. The sample collection is, nevertheless, an important pre-analytical issue in POC testing. The most commonly used sample type in POC testing is capillary blood, although some MPs also analyze venous or arterial blood (37). Several challenges are related to capillary blood sampling. The capillary blood may have a variable composition depending on the blood circulation at the puncture site (37). There are also differences between capillary and venous blood that may influence some analyses, such as hematological and glucose tests. There is a correlation between the depth of the puncture and the amount of blood volume collected, and the correct technique, in addition to puncture aids, can therefore be important to obtain the correct

sample volume (37). Other important preanalytical factors are whether the first blood drop should be analyzed or discharged, which container should be used, and whether the container should be inverted after being filled with blood, which depends on the measurand (37). Processing the samples prior to analysis, for example, using centrifugation, is not an issue when performing capillary blood analysis, and consequently, associated preanalytical problems are eliminated. Inspection of samples in the central laboratory is essential to identify hemolytic, lipemic, and icteric samples, which may cause preanalytical errors. Hemolysis is one of the most common confounding factors in capillary blood sampling and can be triggered if the skin is squeezed during sample collection, therefore, it is essential to use the correct technique (37). Additional possible errors are mislabeling, interchanging, failure to deliver, or misplacement of patient samples (34).

Analytical errors may involve difficulties regarding the MP used, for example, maintenance, the analytical principle, or issues with the reagent lot number (37). Errors in this phase may be difficult for physicians to recognize unless the results are incredibly divergent from what is expected, and the erroneous result is therefore likely to be used in the diagnosis and treatment of patients. Depending on the MPs used, some MPs have quality control processes integrated from the manufacturer to reduce errors and potential hazards. Such quality control processes can detect reuse of test strips, and additionally, some new glucose meters also test for strip damage, humidity, and other hazards that might influence the test results (37). There are concerns about the effect of using expired reagents on analytical quality in POC testing (38). Therefore, there are recommendations to check and discharge any reagents with expired dates, since the chemical coating of the test strips breaks down as the reagent ages, which in turn can lead to biased test results (39). Degeneration of test strips does not only happen after the expiration date, but it may also occur after a certain time interval after the containers containing the testing strips are opened (37). However, some POC MPs use barcodes that identify each bottle within a lot and start a countdown to the expiration date when a bottle is first scanned (37). Similarly, some POC MPs have calibration factors that have been integrated into the product by the manufacturer, so the POC MP initiates an automatic calibration as the test strip or cartilage is scanned. There are some MPs, for

example, glucometers, that possess lockout functions, which prevent MPs from being used in cases where quality control testing has not been documented (37). The lockout function can also be used to ensure that only qualified operators with valid training are handling the POC MPs by requiring that the operator's ID is scanned before analysis (37). In cases where there is an electronic transfer of patients' results from POC MP to medical records, the entry of incorrect patient IDs can impede the transfer of test results. Since the test results are available on the POC MP, treatment can be initiated without the results being documented in the patients' records, and vice versa. An incorrect patient ID entered into a POC MP can result in test results being transferred to the wrong patient record. Positive patient IDs, where a second identifier is mandatory, are integrated into many new POC MPs and ensure that patients' information is correct before testing is allowed (37).

Errors in the postanalytical phase may not be specific to POC testing but are relevant to obtaining high-quality test results in POC testing. Errors may occur in the validation of test results, especially in situations where the tests are performed under high pressure to fulfill short turnaround times. There is also the possibility of incorrect or delayed reporting of results, erroneous entry of test results into mismatched patient records, or a completely missed notification of test results (34). Errors in the postanalytical phase are often related to the organizational settings, for example, how well the POC MPs, laboratories, and wards are linked. Additionally, the number of controls in POC testing is often limited compared to the central laboratories, and as a result, there is an increased risk that errors can impact patients (40).

1.3 Analytical quality

All laboratory analyses are subject to uncertainty in the analytical phase, and regardless of how carefully the analysis is performed, every single measurement differs slightly. Analytical variation can either be random or systematic and will be discussed further below. Additionally, test results are influenced by uncertainty due to biological

variation. Biological variation is a natural variation that occurs either within a person (known as intra-individual biological variation, which is the biological variation around a homeostatic set-point) or between persons (called inter-individual biological variation, which is the variation between individuals with different homeostatic set-points) (41). Typically, analytical variation should not exceed more than half of the intraindividual biological variation. This is because when the analytical variation exceeds more than half the intraindividual biological variation, it constitutes less than 12% of the total variation (analytical and intraindividual biological variation). It is, however, not always possible to achieve an analytical variation of less than 12%, and the expected analytical variation for different analyses is summarized in different manuals (42). The biological variation for different analyses can be obtained from the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Biological Variation Database (43).

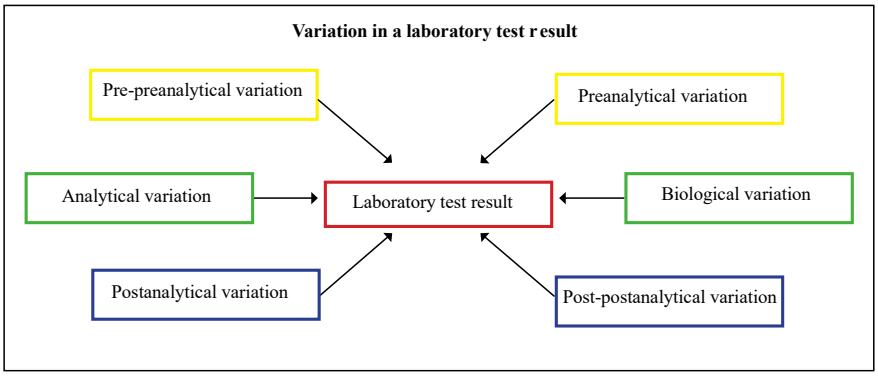


Figure 2. Variation in a laboratory test result. The figure illustrates different forms of variations influencing the laboratory test result.

Random errors are part of analytical variation, and they limit the precision of the performance of the MPs. The precision can be quantitatively expressed as imprecision, also referred to as analytical variation. Imprecision can be stated numerically as the standard deviation (SD) or the coefficient of variation (CV), assuming the MP is performing under given specifications. These specific conditions (listed with an

increasing degree of variability) are repeatability, intermediate precision, and reproducibility (1). These conditions may also be divided into imprecision within one laboratory (comprising repeatability and intermediate precision) and imprecision between laboratories (reproducibility).

The calculated SDs and CVs become more reliable as more data are included, however, there is always a level of uncertainty associated with calculations, characterized as confidence intervals (CI). The CI denotes the uncertainty of a performed measurement, and the prediction interval (PI) denotes the limits where a future value from the same population will fall with a given probability. The limits of the PI are wider than the corresponding CI, but the intervals approach each other as *n* increases (2). When repeated measurements are performed on the same aliquot, the dispersion of the results will, in general, follow a normal distribution scattered around the mean. The expected imprecision of the measurement results gives a statistical probability of 68.3% of the results falling within ±1 SD, 95.4% within ±2 SD, and 99.7% within ±3 SD of the mean (3).

In contrast to random error, which may have both a positive and a negative direction from the mean, systematic error (bias) is caused by the deviation of all the measurement results in one direction. A difference between the observed mean and the expected result is then introduced. Bias can be expressed as either a percentage or an absolute value. Calculating bias is, however, often more challenging than calculating imprecision, as the true value may not be known. Ideally, the measurement results should be compared to a traceable reference MP to assess whether systematic bias is present. Otherwise, an indication of the MP's performance can also be obtained through participation in EQA or by the participants' internal quality control (IQC), demonstrating a shift of the results in one direction. Trueness is inversely related to bias and characterizes how well an MP is calibrated to a traceable reference system.

An MP's imprecision and bias make up the total error (TE) of a measurement result. The TE gives an estimate of the limits of an interval, whereby the unknown "true" value can be found with a defined probability (usually 95%). The TE is useful both for

evaluating a laboratory's assay quality and for setting quality goals for different clinical tests (3). Different approaches have been suggested for calculating TE, and the CLSI's recommendations are elaborated in the EP21A guidelines (44). The TE is expressed in qualitative terms as accuracy.

After calculating the TE limits, the allowable TE (TE_a) for a particular analysis is determined. TE_a refers to the amount of error that is allowed for the particular analysis. The TE must not exceed the limits of TE_a to ensure that the MPs are of adequate analytical quality (2).

An objective criterion for determining TE_a, and thereby whether the analytical quality is acceptable, is essential. TE_a can be estimated by establishing analytical performance specifications (APS) using three different models arranged hierarchically (45, 46). In the first model, the APS are based on the effect of analytical performance on clinical outcomes. This effect can be determined using direct outcome studies, ideally using randomized controlled trials, which investigate how a test's analytical performance impacts clinical outcomes (Model 1a). A more pragmatic alternative is utilizing indirect outcome studies, for example, simulation or decision analysis, where the influence of a test's analytical performance on clinical classifications or decisions (and consequently the probability of clinical outcomes) is investigated (Model 1b). This approach is often used when establishing APS in laboratory practical guidelines (47, 48). Although Model 1 is preferred, it can only be used in situations where the links between the test, clinical decision making, and the clinical outcome are straightforward and measurable (49). The second model defines APS based on biological variation. The approach attempts to minimize the ratio of the analytical variation to the biological variation, which will identify MP performance relating to medical requirements. The data for this approach are acquired from studies of biological variation. In the third model, the APS are based on the "state of the art" in the field. This model is benchmarked using the highest level of analytical performance technically achievable, or, in other words, the analytical performance obtained by a particular percentage of laboratories. There is no official consensus on how to determine specifications using

this approach, but data derived from EQA schemes or empirical data have been suggested (2).

1.4 Analytical quality control

Several quality control systems can be implemented to detect potential errors. In 2000, the Swiss cheese model of errors was introduced, which uses the metaphor of stacked slices of cheese to describe the system. In this metaphor, the defense layers of the system are like slices of cheese, and the weaknesses in the system are represented by the holes (50). Although no system is entirely safe, since there are holes in each defense layer that allow errors to pass, multiple layers will lead to most errors being caught.

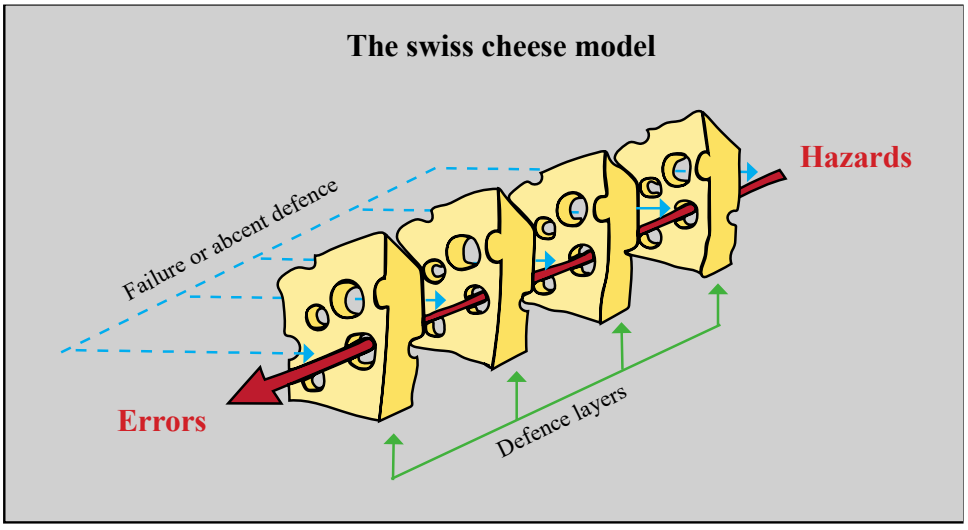


Figure 3. The swiss cheese model. The figure illustrates that several defense layers lie between the hazards and errors, but flaws in each layer may cause errors to happen. The figure is modified from J. Reason figure "Swiss cheese model" (50).

A quality management system is constructed to reduce laboratory errors and ensure that the test results are suitable for supporting decisions about patient care. The quality

control of the analytical process consists of IQC and EQA, and the results obtained from patient sample testing can also be used to monitor performance.

1.4.1 Internal quality control

IQC is continuous, day-to-day surveillance of the stability of the laboratory's MPs and working routines. This evaluation aims to identify, decrease, and correct any insufficiencies in the laboratory's analytical process and to verify that the MPs meet the predetermined quality specification at the time of the analysis. This fundamental characteristic of IQC, coupled with an evaluation of analytical performance, makes it possible to intervene without delay if any insufficiency is detected (3).

In the central laboratory, the IQC material of at least two different concentrations is measured along with patient samples (51). The target values and the SD for the IQC materials are calculated, and statistical control rules are established based on the TE_a and APS for the MPs. The control rules are statistically set up to alert the user if the IQC's TE exceeds the TE_a, which indicates that some patient samples analyzed in the run might have a TE that exceeds TE_a, and there is a risk of erroneous patient results (3).

IQC in POC testing may present further challenges. The POC MPs are often operated by persons who do not hold a laboratory degree and need to be convinced that IQC is important to detect errors. Unfortunately, the evidence that measuring IQC improves POC quality is limited (52, 53). The MPs utilize different technologies that require different levels of IQC. MPs using technology similar to the wet chemistry in central laboratories, for example, require more frequent IQC than cartilage- or strip-based MPs, where the chemical reaction takes place inside the cartilage/strips. Additionally, some MPs have "built-in" controls that can either be placed in the MP itself to assess whether the electronics of the MP are working properly, or they can be placed in the cartridges or strips. However, it has been argued that relying only on such control systems is inadequate, and IQC is recommended (54).

1.4.2 External quality assurance

EQA entails continuous monitoring of the analytical quality of laboratory analysis, as well as confirmation that the test results conform to the expectations required for patient care (55). Additionally, several laboratories perform EQA as part of a framework that is required when seeking accreditation or to assess pre- and postanalytical efforts (56).

ISO 17048 provides information on how EQA programs should be organized and managed by the laboratories (57). Generally, the EQA surveys are performed by an external organization circulating sets of quality control samples, the EQA material, to its participants. The EQA material simulates clinical samples with unknown concentrations and should be analyzed in the same manner as patient samples. The results of the analyzed EQA material are reported to the EQA organization for evaluation. The EQA organizer assigns target values for the EQA material, and the participants' analytical quality is evaluated by assessing the difference between the participants' results and the target values. A report is prepared that includes the participants' results, the MP used, the expected target value, and an evaluation of the participants' analytical quality. The EQA organizer also offers group data analysis at a national level, where the results from different participants using the same MPs for the same measurand are compared. Verification of whether the participants' performance and MPs are in conformance with other participants is then possible (55).

The nature of the EQA material is highly important, since it affects how the results obtained from the EQA scheme can be interpreted (58). Information on whether the EQA material is commutable or likely to be commutable, if the material is known not to be commutable, if there are any specific limitations in the commutability of the material, or if the commutability has not yet been assessed, should be specified for each survey (58). When distributing a commutable EQA material, the samples mimic real clinical patient samples as closely as possible, and the participants' results reflect those expected from measuring real patient samples (55). However, providing commutable

EQA material for all measurands and all MPs is challenging, and in such cases, noncommutable EQA material is used in the EQA programs. Noncommutable EQA materials do not behave like clinical patient samples and therefore only allow for evaluation of the participants' performance in relation to the MP used (55). The EQA material and commutability are discussed in detail later.

A target value and its associated uncertainty must be established by the EQA organizer to enable interpretation of the EQA results. Ideally, the target value is obtained from a reference MP or a comparative MP with high specificity, traceable to a reference MP (55). The target value can also be established by value transfer using a commutable certified reference material (CRM) (59, 60). Preferably, the traceability should be determined by reference MPs and CRMs listed in the database of the Joint Committee for Traceability in Laboratory Medicine (61). There is also an option for weighted values in cases where the purity of the material, the accuracy of the MP, and the similarity between the purified material and its endogenous form in human samples are verified (55). Establishing a target value using a reference system is most beneficial in cases where the EQA material is commutable. When the commutability of the material is not known, it is not possible to know whether deviations from the target value are due to a calibration bias or a matrix-related bias. If a reference system is not in place, the mean or median of all participant results can be used as a target after excluding outliers, providing that the EQA material is commutable, and the same result is expected for all MPs used. In cases where the commutability of the EQA material is not established, it is common to separate the participant results into "peer groups" of MPs with similar technology, which are assumed to obtain equivalent results when measuring the EQA material after excluding outliers (3). Usually, the peer group consists of MPs from the same manufacturer. How the peer group target values are calculated is important, attention must be paid to how outliers are treated, and the amount of uncertainty associated with the target value, especially in situations where the data are limited. An all-participant mean/median can be used in situations where all MPs have a similar matrix-related bias or are equal in size, as there otherwise is a risk of the target mainly reflecting the largest MP group (55).

The difference between the EQA results and the target value is evaluated to assess whether the participants' analytical quality is adequate. The EQA organizer creates the APS for the measurands involved to determine whether the deviation from the target value is acceptable. The APS are often denoted as units or percentages based on the target value, determining the lower and upper acceptance limits (58). However, the criteria for evaluating adequate performance differ among EQA providers (62). Some countries have regulations defining the criteria that the EQA organization is required to follow, while other EQA providers assign evaluation criteria (the APS) themselves (63). Some EQA providers use statistical limits, for example, z-scores, where there is an assumption that the EQA results of the MPs are acceptable if they are in concordance with others using similar MPs (3). Other EQA organizers establish their APS based on clinical criteria or biological variation (64, 65). Consequently, there are large differences in the APS for the same measurands used by different EQA organizers due to, for example, differences in the data on which the APS are founded or the different criteria used in establishing the APS (58). There are, nevertheless, published recommendations on different models to use when establishing APS, along with guidance on how to achieve harmonization among different APS for the same measurands (58, 66). The APS used for EQA materials often have wider limits than the corresponding TE for clinical patient samples, as both the variability for the particular measurand and the error components unique for EQA materials must be taken into account. Examples of such unique error components include between laboratory variation in the calibration of MPs, matrix-related bias due to differences in the lots in peer groups, the uncertainty of the assigned target value, as well as stability and homogeneity issues in the EQA material. When the APS are established for a single measurement, the limits are wider since TE is assessed. When multiple EQA results have been retrieved from the EQA material, bias and imprecision can be evaluated separately, and the APS limits can be narrower (58). In cases where the EQA material is noncommutable and the target value is assigned by calculating the peer group's mean/median for similar MPs, the APS should consequently be stricter, as the variation in the results is limited to the particular MP group (58).

Interpreting EQA results when the EQA material is commutable has the advantage that the relationship between the EQA results corresponds to the relationship observed for clinical samples (3). The participants can then determine the accuracy of their clinical patient results by comparing them to the EQA results, with a target value preferably established by a reference MP. The results should then, in theory, be the same for all MPs, reagent lots, and routine MP calibrators, and the variation seen in the results reflects real differences between the MPs, lots, or calibrators (55). Consequently, the results from EQAs can be used to assess the harmonization status of patient results between different MPs, and the outcome of calibration standardization or harmonization among participants can be assessed (67-69). MPs that do not perform as expected can be identified, and the calibration can be corrected by the manufacturer. The EQA results can also be useful for detecting measurands that need improvement and, hence, enable the initiation of adjustments to promote reaching clinical requirements (70, 71).

Interpreting EQA results when the EQA material has unknown commutability has limitations because the matrix in noncommutable EQA materials can induce bias between MPs, which is not seen when measuring clinical patient samples (67, 72-74). Direct verification of the accuracy of clinical patient results based on the EQA results is, therefore, not possible since actual differences cannot be separated from any matrix-related bias. Still, quality can be assessed, and participant performance can be evaluated. It is also possible to verify that the participants are using the MPs in accordance with the manufacturer's instructions and the other participants in the same peer group, for example, using similar MPs (55).

EQA programs can be assigned to one of six categories based on their ability to evaluate the participants' performance (Table 1) (3). The EQA programs' ability to evaluate performance depends on the EQA materials' commutability, the process used to establish the target value, and whether the participants analyze the EQA material in replicate (55). Category 1 is preferred, as it includes commutable EQA material and target values that are traceable to a reference MP or CRM, which allows for assessing both participant performance and the MPs for calibration traceability, reproducibility,

repeatability, and harmonization among participants and MPs. In category 2, the EQA material is analyzed in singlicate, and consequently, the within-laboratory reproducibility cannot be assessed, otherwise, the category has the same qualities as category 1. Categories 3 and 4 include EQA programs using commutable EQA material, but there is no reference measurement system in place for establishing the target value, which limits the ability of the evaluation to assess the uniformity, or harmonization, among participants and MPs. EQA programs in categories 5 and 6 use noncommutable EQA materials, and the evaluation is limited to a peer-group comparison of the EQA results (55).

Table 1. EQA scheme design. Evaluation capability of different categories of EQA schemes. The table is modified from Miller et al. (3).

EQA scheme design										
EQA category	**EQA material characteristics**			**Evaluation capability**						
	Commutable	Reference target value	Replicate measurements	Trueness			Imprecision		Standardization Harmonization	
				RMP/CRM	Overall	Peer group	Within-lab	Between-lab	RMP/CRM	Mean lab
1	Yes	Yes	Yes	X	X	X	X	X	X	X
2	Yes	Yes	No	X	X	X		X	X	X
3	Yes	No	Yes		X	X	X	X		X
4	Yes	No	No		X	X		X		X
5	No	No	Yes			X	X	X		
6	No	No	No			X		X		

1.4.3 EQA in primary care in Norway

EQA in Norway is provided by the Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus). More than 3600 participants are voluntarily engaged, and GP offices and nursing homes account for the majority of participants (over 99% and 96% of all GP offices and nursing homes are enrolled, respectively) (75). Currently, Noklus distributes EQA schemes for 21 analyses, each of which is circulated between one and four times annually, engaging between 80 and 2900

participants (75). Noklus is one of the few organizations globally in possession of EQA data covering nearly 30 years of primary health care (76).

Noklus provides a quality management system covering the total testing process, complying with the requirements of the NS-EN ISO 9001:2015 and the NS-EN ISO/IEC 17043:2012 standards (57, 77). Figure 4 illustrates how Noklus works with the participants to improve the quality management system and, in the end, reach the common goal of providing safe patient care. Initially, the process starts with Noklus producing "in-house" EQA materials resembling real patient samples, which are distributed to the participants by mail. The participants analyze the EQA material in duplicate on two different days. Analytical results, information on the MP used, and additional characteristics of the participants are reported back to Noklus. At Noklus a target value is assigned for some measurands by a reference MP or value transfer using a commutable CRM, and for others using a peer group median after excluding outliers. When using a target value from a reference MP, Noklus obtains the value from an external body. When using CRMs, the target value is obtained by measuring the EQA material and the CRMs using an MP with excellent precision and then calculating an inverse regression line against which the EQA material can be assessed. In cases where a peer group target value is used, the target value is calculated according to ISO 13528:2015 (78). At least eight participants are needed in a peer group for the estimated target value to be statistically robust. An interval of uncertainty around the target value is established both to address the uncertainty of the defined target value and to ensure that the APS are relatively wider at lower concentrations of the measurand. The target value and its uncertainty are communicated to the participants.

Noklus uses APS, which are defined using a combination of fixed percentage limits, intraindividual biological variation for the particular measurand, state-of-the-art and expert opinions to assess the participants' analytical quality (58). The participants receive an evaluation of the trueness and precision of their duplicate results graded as "very good", "acceptable", or "poor" based on the APS. When the data analysis is performed and the test results are interpreted, Noklus prepares an individual feedback report for each participant.

Through the EQA, the participants can communicate any difficulties they experience when performing the EQA back to Noklus. Noklus supports the participants in the process by answering questions and providing advice to enable them to achieve the best laboratory quality possible. Noklus also provides three preanalytical questionnaires and one postanalytical questionnaire yearly to the participants to ensure that the total testing process is covered.

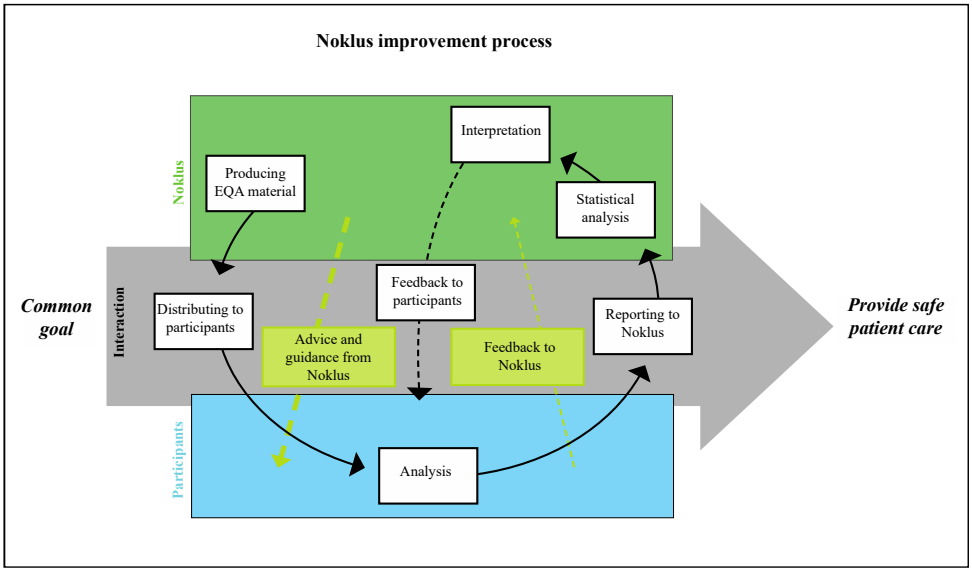


Figure 4. Noklus improvement process. An illustration of how Noklus works to continuously improve quality. Noklus and the participants have a common goal of providing safe patient care. The green area illustrates the contribution from Noklus, the blue area illustrates the contribution from the participants and the grey arrow illustrates the interaction between the two.

1.5 Measurement procedure performance

Obtaining clinical test results that are comparable irrespective of the MP used is a high priority for the utility of laboratory information. While standardization refers to MP results being traceable via calibration to any higher-order reference material calibrator,

ideally to the Système Internationale (SI), harmonization denotes that the MP results must be equivalent without emphasizing the process of achieving such results (3).

Harmonization and standardization aim to provide consistent and comparable laboratory results regardless of the MP used and the time and place of the analysis (79). They allow for common reference intervals and the use of clinical practice guidelines with fixed decision limits. This also enables laboratory data to be gathered from various sources and thereby helps to identify public health issues, monitor health programs, and facilitate comparisons of data derived from different studies (80). If the EQA material is commutable, standardization allows for common target values for EQA organizations, which can be used to objectively evaluate the performance of IVD MPs and to provide the participating laboratories with an accuracy-based common grading of their performance (81).

The entire test process, i.e., the preanalytical, analytical, and postanalytical aspects, must be taken into consideration to achieve harmonization.

Preanalytical aspects include having a standardized terminology to make sure the correct test is ordered, assuring that the patients are following the rules of preparation, and that the specimen collection process is performed without errors.

Postanalytical aspects involve using standardized terminology to make sure the test results are correctly communicated, ensuring that the reporting units are identical to avoid any misinterpretation, and providing uniform reference intervals and decision values to prevent different clinical interpretations of the same laboratory results.

The pre- and postanalytical aspects of harmonization will not be further discussed in this thesis.

Analytical aspects to consider are the measurand, traceability, and commutability.

The measurand is defined as the quantity intended to be measured (1). Inadequate definition of the analyte can lead to different entities being measured by different MPs. Human chorionic gonadotropin, for example, is known to have several isoforms

depending on the clinical setting, for example, tumors and pregnancy (82). Another prerequisite is the analytical specificity of the MPs. If the MPs are influenced by other substances in the sample, different results may be obtained from the same sample (83).

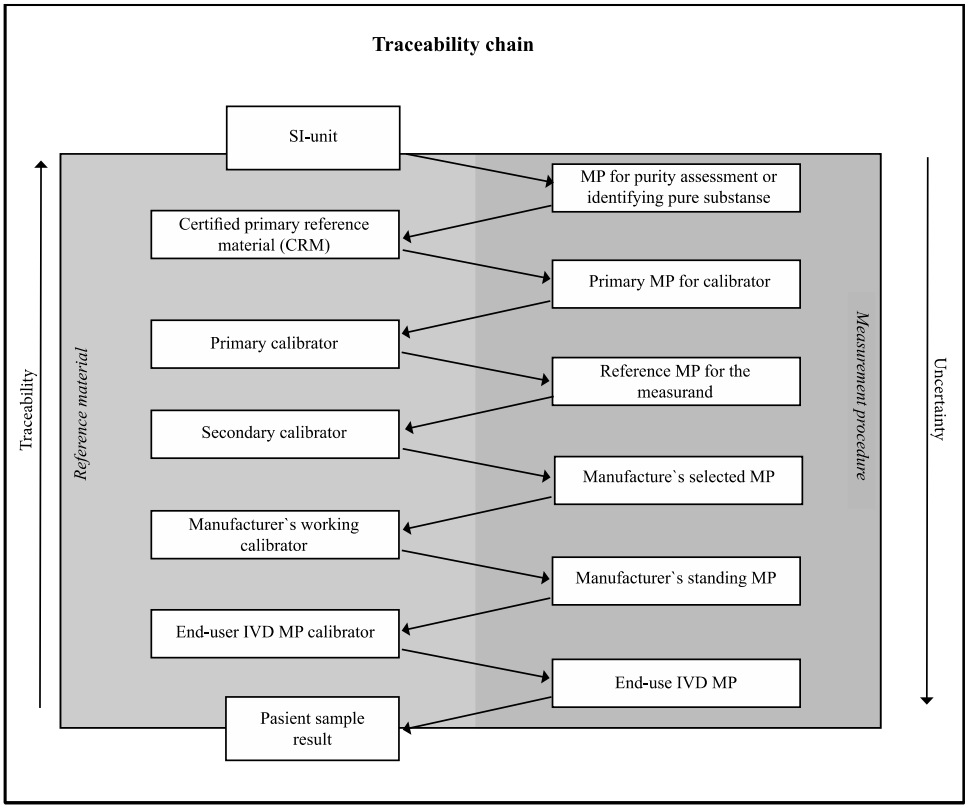


Figure 5. Traceability chain. An illustration of the traceability chain starting with the SI-unit at the top of the hierarchy and ending with the patient result. Traceability is achieved by value assignment through an unbroken chain of calibrators. Since all calibrations have stated uncertainties, the amount increases the further down we are in the traceability chain. The figure is modified from ISO 17511:2020 (79).

Traceability aims to link clinical results from routine MPs to higher-order reference standards (79). This is achieved by establishing a hierarchical reference measurement system consisting of reference materials and reference MPs, where each level is joined through an unbroken chain of calibrators.

The reference MPs and reference materials are the key components of the traceability chain. The SI is at the highest hierarchical level of the chain, followed by a primary reference material and a primary MP. However, in cases where this is not possible, it may be possible to establish traceability to the WHO International Biological Preparations. This is an "artefact standard" reporting concentrations using the International System of Units and is used, for example, in assessing total Hb (80). Ideally, a primary MP is used in the value assignment of the reference materials, however, (for various reasons) not all analytes are compliant with the primary MPs available. The primary reference material (often several concentrations) is used as a calibrator for a secondary reference MP. The secondary reference material is used by the manufacturers to calibrate the internal MP they have selected to assign values to their working calibrator, and this is also known as the master lot of the calibrator. A master lot is used by the manufacturer to calibrate the manufacturer's standing procedure, which is used to assign values to product calibrators. The master lot can be used over several years to assure traceability of different lots of product calibrators, which in turn are used to calibrate the routine MPs in clinical laboratories used to measure patient samples.

Each step in the traceability chain is associated with some degree of uncertainty, which, in combination, represents the total uncertainty of the final patient result. Approaches to reduce imprecision in each calibration are therefore essential, in addition to adopting the minimum number of traceability steps necessary (84).

Traceability of calibrators and control materials to reference materials and/or reference MPs is required according to the EU Directive on IVD Medical Devices. In a database held by Bureau International des Poids et Mesures (JCTLM), information can be obtained for all certified reference materials, reference MPs, and reference laboratories complying with ISO standard 17511 for reference systems for measurands (61, 79).

There are several measurands for which no higher-order reference materials or MPs are available. In these cases, traceability ends with the manufacturer's master lot, and consequently, the results from patient samples will differ depending on the MP used at

the laboratories. However, there are ongoing discussions about making calibrations traceable to harmonization protocols. These protocols are based on international consensus and aim to obtain uniform measurements for MPs that lack either a reference material or a reference MP (85).

Commutability of the reference materials used as calibrators in the traceability chain is essential to ensure that they are fit for use. The noncommutability of reference materials used as calibrators in the traceability chain is a contributor leading to a lack of agreement among different MPs (86). Commutability is especially important for secondary reference materials that are used as common calibrators and may not be encountered until after the production and certification process, since the reference MPs involved are relatively matrix independent (84, 87). Nevertheless, commutability can be a challenge due to alternations in the matrix during the preparation of the reference material. Manufacturers often provide target values for calibrators intended for specific MPs, which can be used to correct for any noncommutability and make patient sample results from this specific MP traceable. However, the calibrators will not be commutable when used with different MPs.

1.5.1 Standardization of the CRP, glucose, and Hb measurement results

CRP

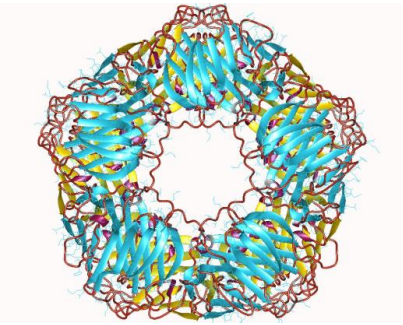


Figure 6. CRP. Human C-reactive protein by Ramadan, M.A., Shrive, A.K., Holden, D., Myles, D.A., Volanakis, J.E., DeLucas, L.J., Greenhough, T.J.; User: Astrojan (https://commons.wikimedia.org/wiki/File:1lj7.jpg). Licenced under Creative Commons BY 3.0.

There has been a great effort since the late 1960s to standardize serum protein measurements to reduce the poor between-laboratory agreement seen for serum protein measurements results (88). The introduction of the ERM-DA470 (CRM-470) has especially resulted in improved harmonization of the serum protein measurement results in general (89, 90). For CRP, this standardization has permitted the use of common reference intervals and thereby allowed for patient results to be compared over time between different laboratories.

Standardization of serum proteins is challenging due to genetic variability, post-translational alternations, ligand-binding, proteolysis, and the fact that many proteins can differ substantially in their concentration in serum. Native CRP is a cyclic pentameric protein with an annular configuration. The molecule consists of five non-glycosylated subunits, each made up of 206 amino acids (91). The subunits, known as protomers, contain a pocket with two bound calcium ions. The calcium ions are crucial for all ligand binding of CRP and to enable the molecule to retain its pentamer structure

(92). There have been indications that the CRP molecule may show heterogeneity in vivo, nevertheless, no polymorphism has yet been confirmed (92).

Currently, there is no established reference MP for analyzing CRP stated by JCTLM, and the standardization process therefore relies on the availability of reference materials (61, 93). The CRP MPs are traceable to the certified reference materials ERM-DA474/IFCC (CRP in processed human serum) and ERM-DA472, which are traceable to SI via ERM-DA470/IFCC and CRP WHO 85/506 (94, 95). The reference material for CRP is a separately produced liquid frozen material, and CRP is not part of the common reference material for serum proteins, ERM-DA470k/IFCC. The ERM-DA470k/IFCC was produced to fully replace the common reference material for serum proteins DA470/IFCC. However, the newly produced reference material for serum proteins was shown to be unsuitable for CRP. The lyophilization process that takes place to create the material may have altered the pentameric structure of CRP, thereby increasing the amount of monomeric CRP. Since monomeric CRP in vivo is predominantly membrane-bound and is, as a result, not measured by immunoassay methods, a liquid frozen nonlyophilized reference material was prepared instead (93).

Commutability of the ERM/DA/IFCC material has been evaluated for several batches using different buffers, antibodies, and reagents. The CRM was assessed as commutable for most immunoassays, however, if the CRM is intended to be used for calibration, commutability should be verified for the particular MP involved (94).

Glucose

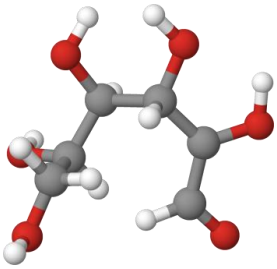


Figure 7. Glucose. "Druesukker" by Johannes Botne (https://snl.no/druesukker). Licensed under Creative Commons BY SA 3.0.

The measurement of glucose is used in a variety of clinical settings, with a range of different sample types and measuring techniques being used. Harmonization of glucose measurement results has therefore been challenging, but increased attention from, for example, the International Federation of Clinical Chemistry (IFCC) Scientific Division's Working Group on Selective Electrodes and Point of Care Testing (IFCC-SD-WG-SEPOCT) has been conducive. Both glucose measured in serum and whole blood have adequate harmonization/standardization according to the International Consortium for Harmonization of Clinical Laboratory Results (85).

Glucose is a small measurand that is well-characterized in both chemical and physical terms (80). The molecule is a hexose monosaccharide existing in two different isomers, D-glucose and L-glucose. The two isomers mirror each other but are otherwise structurally identical. D-glucose is biologically active and occurs in two forms, α-D-glucose and β-D-glucose, which make up 36% and 64% of D-glucose at equilibrium, respectively (96).

In whole blood, glucose is distributed between erythrocytes and plasma, and although the molality of glucose is equal throughout the sample, the concentration of glucose is higher in plasma compared to the intracellular glucose concentration. Samples can be

whole blood (venous, arterial, or capillary), plasma, or serum samples from centrifugated venous whole blood. The glucose measurements are performed using various types of MPs with different principles for detecting glucose. The units for reporting glucose measurement may also differ depending on the laboratory, whereas many now prefer mmol/L, some still report glucose values in mg/L. To harmonize glucose measurement results, the IFCC-SD-WG-SEPOCT recommends reporting glucose concentrations in plasma irrespective of sample type or MP used, using the unit mmol/L and applying a conversion factor of 1.11 to make whole blood and plasma results equivalent (97). Another variable to take into consideration when analyzing hemolyzed whole blood samples is hematocrit. A patient with low hematocrit (< 30%) can have falsely increased glucose results, while a patient with high hematocrit (> 60%) can have falsely reduced glucose results. The degree to which hematocrit affects the measurement results depends on the MP used and must be taken into consideration.

Despite the challenges regarding the harmonization of glucose measurements, the measurand has full traceability to a complete reference system. The SI unit of glucose (mmol per liter) is included in the primary reference material, NIST SRM 917b (crystalline glucose). This material is composed of chemically purified glucose, and the material is used to prepare calibrators used for isotope dilution mass spectrometry (ID GC-MS), which is the established secondary reference MP for glucose (61, 98, 99). ID GC-MS combines mass spectrometry with isotope dilution and is currently the most accurate known method for analyzing glucose. The ID GC-MS reference method is suitable for the analysis of glucose in whole blood (Deutsche Vereinte Gesellschaft für Klinische Chemie und Laboratoriumsmedizin, Germany) and for analysis in serum (Gent University, Belgium) (61). ID GC-MS is used to measure glucose concentration in the CRM SMR 965b (glucose in frozen human serum) (61).

Commutability of the secondary reference material NIST SMR 965b has not yet been evaluated. The certificate of analysis states that although the National Institute of Standards & Technology (NIST) is unaware of any problems with commutability, the reference material might not be commutable with all routine glucose MPs (100).

Hb

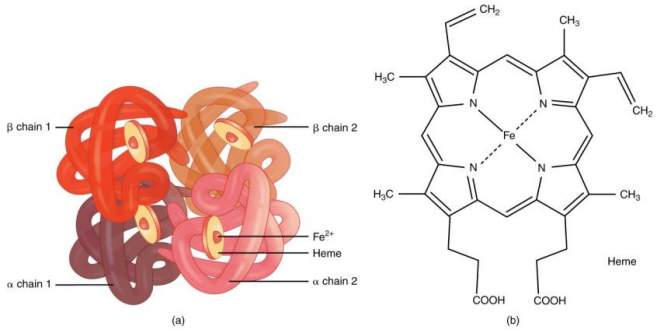


Figure 8. Hb. Hemoglobin by OpenStax (https://openstax.org/books/anatomy-and-physiology/pages/1-introduction). Licenced under Creative Commons BY 4.0.

The standardization of hemoglobin (Hb) began in the early 1960s after widely different results for Hb were obtained using a variety of different MPs (101). A consensus on using photometric determination after the conversion of Hb into cyanmethemoglobin (HiCN) was established by the International Council for Standardization in Hematology (ICSH) (101). HiCN is the current reference MP for Hb measurements stated by JCTLM, and the harmonization status of Hb is referred to as adequate (61, 85).

The measurand Hb is structurally and physically well-defined. The molecule is made up of a protein part, globin, surrounded by four hem molecules forming a tetramer. The globin consists of two linked pairs of polypeptide chains. Approximately 97% of the total Hb in adults is HbA, which has 2 α- and 2 β-chains, the remaining Hb are HbA_2 composed of 2 α- and 2 δ-chains and some fetal Hb, HbF, consisting of 2 α- and 2 γ-chains. Each of the polypeptide chains is linked to hemoproteins, which are circular porphyrins s (a class of water-soluble, nitrogenous pigments) with an iron atom (Fe^{2+}) attached (102). There are several different Hb derivates in vivo, with different absorption spectra present in the blood. As a result, prior to spectrophotometric

analysis, whole blood samples must be diluted to obtain a homogenous solution. The transformation of Hb into hemiglobincyanide before analysis is at present considered to be the most reliable procedure (103).

Hb measurements are not traceable to SI but are traceable to a conventional reference procedure and a conventional reference material defined through consensus by the ICSH (80). The reference MP stated by JCTLM is HiCN, which measures the color product of hemiglobincyanide, facilitating a spectrophotometric method (61). The method is accurate, reproducible, and thoroughly investigated, and the reagents are inexpensive. HiCN is calibrated using BCR-522 bovine blood lysate (hemiglobincyanide) from the Community Bureau of Reference (104). This reference material is intended to calibrate the HiCN to validate the secondary reference materials. At Noklus, participants have their Hb EQA results assessed against a target value obtained from a HiCN method at the Odense University Hospital (Odense, Denmark).

1.6 The EQA material

Since the EQA material is the key component in the EQA scheme, it should ideally possess characteristics enabling the assessment of individual MPs and their accuracy, precision, and calibration traceability, as well as a collective evaluation of the laboratories, assessing their reproducibility, standardization, and overall harmonization (3). The EQA material must possess a range of qualities to be able to fulfill these requirements. First, the appropriate sample type (e.g., whole blood, serum, or urine) must be available in a sufficient amount, the material should be at a clinically relevant level, and should cover clinical decision points of the measurand in both normal and pathological concentrations. The material must be practical to handle, free of any pathogens (unless they are the purpose of the EQA scheme) and should be available at a reasonable price. It should be homogenous across aliquots and stable for both short and long times under the conditions in which the material will be transported and stored. Finally, the EQA material must behave in the same way in the routine MPs as

the clinical samples they mimic, hence, they must be commutable. Commutability between the EQA material and clinical samples is essential and affects both the EQA scheme design and the ability to interpret the results obtained from the participants (55). This will be discussed in more detail below.

The EQA materials distributed by the EQA organizer are intended to be used as surrogates for authentic clinical samples usually measured at the laboratory. They must therefore mimic clinical samples as closely as possible, and the EQA materials should (ideally) be directly entered into the laboratory workflow without any special handling.

1.6.1 The EQA material for POC CRP, glucose, and Hb

POC testing is designed for rapid analysis. Using whole blood, which is easy to obtain from a capillary finger prick, does not require extensive processing like centrifugation and makes the sample type ideal for many laboratories in primary health care. Another positive feature of using whole blood capillary samples is that the total sample volume needed for testing is minimal. The POC MPs for CRP, glucose, and Hb are usually calibrated to measure whole blood.

Therefore, distributing whole blood EQA materials to the participating laboratories is preferable for mimicking real patient samples as closely as possible. The whole blood material can be prepared from either single donors or pooled samples. The number of aliquots required frequently precludes using single donor samples and, additionally, single donor samples are limited by the possible presence of any interfering substances that may influence the MP and, hence, confound the interpretation of the analytical result. Pooled samples, on the other hand, are usually easy to obtain in the desired amount, and possible interfering substances are most likely diluted. Nevertheless, there is a possibility that substances in the pool might interact and, in turn, modify the matrix of the samples (55).

Noklus obtains whole blood from healthy donors at the blood bank for the EQA material for the joint CRP, glucose, and Hb EQA scheme, and the material is also

pooled to reduce the chance of any interfering substances. However, some of the MPs analyzing glucose have been shown to be incompatible with the use of anticoagulants (ethylenediamine tetraacetic acid (EDTA), lithium heparin, or natrium heparin). Hence, serum EQA samples, also produced at Noklus, are circulated to a low proportion (0.5%) of the participants.

Participants must be provided with samples in concentration ranges that are clinically relevant to ensure that the EQA materials are fit for use. However, using pooled samples from healthy people makes it challenging to obtain the concentration ranges needed for the particular analyte. This is especially true because the EQA materials include three measurands, and each is in two different concentrations. For CRP, which can increase over 100-fold in response to infection or inflammation, as well as for glucose, the concentration of the analytes in the healthy donor samples is too low to cover the clinically important decision points. The Hb concentration, on the other hand, which differs within the normal reference range of healthy individuals, might have an initial concentration that is either too high or too low for the desired level in the EQA material. Analytes can then be added or removed to achieve the preferred concentrations, or it may be possible to dilute the material (55). When spiking the EQA material with a simple and well-characterized analyte in a highly purified form, it is reasonable to believe that the matrix of the material will not be altered (55). However, increased complexity of the added clinical analyte increases the risk of changing the sample matrix and inducing noncommutability (93, 105). The particular analyte can be removed using a highly specific technique, or the material can be diluted using a suspension bearing similar matrix properties to obtain an EQA sample with a lower concentration (55). At Noklus, human native CRP with a purity > 70% (in.vent Diagnostica), and D(+)-glucose monohydrate (VWR) are used to achieve EQA samples with the desired concentrations. The material is either diluted using plasma or erythrocytes are added to obtain both low and high Hb concentrations.

After production, the EQA material is divided into thousands of aliquots to provide each participant with two levels of EQA samples. The occurrence of between-bottle inhomogeneity can be induced, since aliquoting of large pools may take hours. At the

end of the production process, the homogeneity between bottles is evaluated by analyzing the concentration of CRP, Hb, and glucose from several bottles in the entire lot of aliquots produced. Ten bottles are evaluated in total to obtain a reliable assessment of the EQA material's homogeneity, based on the lot size (106).

After aliquoting, the EQA material is distributed to the Noklus participants by mail. Verification of the material's short-term stability during shipment and throughout the analytical period is therefore needed. Noklus has two different approaches for testing the stability of the EQA samples. Using the classical laboratory method, ten EQA samples at each concentration level are exposed to room temperature for two days before being refrigerated for one week (9 days in total). The storage conditions are supposed to mimic the mailing and storage at the participating laboratories during the analytical period of the EQA scheme. The EQA samples, at both levels, are then analyzed ten times each day, and a statistical validation of the EQA material's stability is then performed. The other method of assessing the EQA material's stability is by using the participants' results statistically and assessing whether time is a significant contributor to the variation seen between the samples. In the whole blood EQA material, keeping the glucose concentration stable is particularly challenging. Glycolysis starts immediately after the material is collected, and thereafter, the concentration of the glucose in the material decreases 5–7% per hour if the whole blood sample is left at room temperature (107). In an EQA scheme in which the analytical period lasts over a week, the material needs to be stabilized. The Noklus EQA material therefore has added iodoacetic acids sodium salt and chloramphenicol succinate to maintain the concentration of glucose and control microbial contamination in the samples (78). Adding such stabilizing materials may cause alterations to the sample matrix of the whole blood samples and may result in noncommutability (55).

Making the EQA samples fit for use is a prerequisite, but it is also an intervention that may alter the matrix of the EQA material to such an extent that the samples are no longer representative of authentic patient samples (i.e., they are noncommutable).

1.6.2 Validating commutability

Commutability is one of the most important characteristics of the EQA material. When commutability is not verified, a matrix-related bias may affect the MPs by an unknown magnitude, thereby restricting the EQA organization's ability to evaluate the performance of the MPs and the participants. In the harmonization and standardization process, EQA schemes can be very useful for assessing and monitoring the agreement of laboratory results for the same measurand when using different MPs (108). However, data have indicated that EQA results may not always be useful due to matrix effects within the processed samples (55).

In every chemical analysis, the analyte of interest is set in a compound of other substances, a sample matrix (79). The matrix effects may impact the sample, independent of the analyte, causing matrix-related bias and thereby influencing the analytical result by an unknown magnitude (108). Evaluating the performance of the MPs can be difficult, and in turn can lead to incorrect assessments of the analytical bias, which in turn can affect patient results if the MPs are adjusted based on incorrect data (108). The clinical consequence of matrix-related bias can be misclassification of risk or patients receiving inappropriate therapy (108).

The ISO definition of a matrix is generic and comprises all sample materials, including clinical samples, and all endogenous substances, except the analyte (74). Matrix effects leading to noncommutability of the processed samples are a result of the interplay between many different components in the analytical testing procedure, for example, the design of the MP, the reagents, the sample processing technique, and the processed sample's matrix (108). Matrix effects leading to noncommutability of the processed samples are a result of the interplay between many different components in the analytical testing procedure, for example, the design of the MP, the reagents, the sample processing technique, and the processed sample's matrix (84). However, in practice, differentiating between a lack of specificity and matrix effects can be challenging (74). Interference caused by normally occurring endogenous substances in abnormal concentrations is in most cases not regarded as a matrix effect, however, nonnative forms of an analyte are considered a matrix effect (55).

ISO 17511 defines the commutability between a material and a set of clinical patient samples in relation to a chosen MP (79). The definition thereby refers to commutability as an MP-specific characteristic, hence, the EQA material can be commutable for some MPs, but noncommutable for others. Therefore, whether the EQA material can be used as a common trueness control depends on the number of MPs for which the EQA material is found commutable (55, 74). The causes of noncommutability are, thus, highly dependent on the MP, and the specific reasons leading to matrix effects and noncommutability may be difficult to predict.

There are different consensus procedures in place for validating the commutability of reference materials and processed samples (86, 108). The CLSI has published guidelines in 2014, EP14-A3, *Evaluation of Commutability of Processed Samples Approved Guideline-Third Edition*, and in 2018, the IFCC Working Group on Commutability published three papers with recommendations. The first two papers are relevant for EQA organizations: *IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design*, and *IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples* (86, 109, 110). The establishment of mathematical relationships and assessing whether the results from the processed samples belong to the same distribution as the results obtained from authentic clinical samples when measured on a set of chosen MPs are common to both procedures (86, 108).

The CLSI standard recommends the number and type of clinical samples that must be collected, indicates how they should be handled, how the analysis of the clinical samples and processed samples should be performed, and the number of replicates that is preferable. The CLSI advises the use of Deming regression for data analysis after performing tests to judge the utility of the statistical test for the dataset (outliers, linearity, variance, and homogeneity). The means of the triplicate measurements should then be plotted, and Deming regression should be carried out for each MP combination. A PI is calculated, and an assessment of the processed material can be performed (51).

An advantage of using Deming regression is that the method allows variability for both variables (74).

In 2018, the IFCC launched guidelines for assessing commutability (86, 109, 110). The guidelines are a revised recommendation document in which two different statistical methods are proposed. The first method is recommended for processed materials intended to be used as calibrators, trueness controls, or EQA materials. The second method is suitable only when the processed material is intended to be used as a calibrator (109). The difference between the bias of the processed samples and the average bias of the clinical samples for all pairs of MPs are calculated using the IFCC method for assessing the commutability of an EQA material. Each of the process samples has error bars stating the uncertainty in the bias estimate, which is evaluated against predetermined fixed criteria indicating the maximum allowable bias. A processed material is assessed as commutable when the error bars are entirely within the allowable bias limits. MPs with large random errors will show increased scatter of the clinical samples and larger error bars of the processed materials and are, therefore, more likely to be assessed as noncommutable since the allowable bias limits are the same for all pairs of MPs examined. An evaluation of the MPs included in the commutability study should be done prior to the study to ensure that only MPs with adequate selectivity for the measurand are included.

In general, the conclusions from the commutability study are only applicable to the MPs and reagent lots evaluated in the commutability study, therefore, ideally, all relevant MPs and reagent lots should be included. Nevertheless, since every commutability study in such a case had to be performed several times, depending on the number of reagent lots included, this is not realistically achievable. Both CLSI and IFCC recommendations, however, allow for assuming commutability for subsequent lots of processed materials when they are prepared in the same stringent manner (86).

1.6.3 Validating commutability of the EQA material for CRP, glucose, and Hb

The whole blood prepared at Noklus is exposed to various alterations during production, hence, assessing the material's commutability for the MPs (for which the samples are intended) is of great importance. The EQA material aims to act as a surrogate for authentic clinical samples. Therefore, individual clinical samples that were most likely to contain the matrix that the MPs intended to measure were collected for the commutability study. The number of samples gathered exceeded the minimum requirement, ensuring that there were sufficient data points for the statistical assessment. The clinical samples were selected with the intention of covering the measurement interval of the processed samples, not the total measurement interval of the MPs. This compromise is especially relevant for analytes with large measurement intervals, such as CRP, where obtaining clinical samples covering the entire concentration range is very challenging. The stability of the measurands is an important consideration in the selection of the method used to conduct the commutability study. In the current study, glucose was not sufficiently stable for prolonged storage and had to be analyzed without delay, so the experimental design therefore differed somewhat from the other measurands, since the collection and analysis was spread over multiple days. Noklus aims to include all the participants' MPs in the commutability studies and therefore does not make exclusions based on (for example) precision or selectively of the measurands. However, if the number of MPs must be limited for practical reasons, the market share of the MPs will be the deciding factor. The details of the commutability study will be outlined in more detail in the materials and methods section.

2. Aims

The overall purpose of the thesis was to evaluate the effect of participating in an EQA scheme, and to assess the EQA material used, for evaluating the participants' analytical quality.

Paper I

The aim of the study was determining how the analytical quality of the u- albumin analysis in the GP offices in Norway developed between 1998 and 2012 in relation to the length of the participants' involvement in the EQA and to evaluate what factors in the GP offices were related with good analytical quality.

Paper II

The aim of the study was to evaluate the effect of participation in a POC quality improvement system over time on the analytical quality of CRP, glucose, and Hb, and to determine which factors were associated with good participant performance.

Paper III

The aim of the study was to determine the commutability of the Noklus whole blood EQA material for CRP, glucose, and Hb for the most frequently used POC measurement procedures in primary healthcare in Norway, and to investigate the possibility of using a common EQA target value for each analyte.

3. Materials and methods

3.1 Participants

In paper I, EQA results from GP offices that participated in Noklus and performed the u-albumin measurement were included. The EQA scheme for u-albumin is circulated annually, and from 1998 to 2012, a total of 15 u-albumin EQAs were distributed to the participants.

In paper II, EQA results from participants conducting either the CRP, glucose, or Hb measurements were included, i.e., GP offices, nursing homes, emergency primary-healthcare facilities, family healthcare centers, and providers of home care and occupational healthcare. In some of the statistical calculations, home care and family healthcare centers were grouped as "others" since the groups contained relatively few participants. The analytes CRP, glucose, and Hb are covered by one EQA scheme distributed to the participants biannually. EQA results from 2006 to 2015, in total, 19 EQAs, were included in the study.

The following inclusion criteria were established for papers I and II. The specific MPs had to be used by more than eight participants, and the participants had to provide numerical results (not only a comment) to the EQA scheme. The participants who used quantitative MPs had to report results for both replicate measurements at both levels (four analytical results). In paper I, both semi-quantitative and quantitative MPs were included, and in paper II, all MPs involved were quantitative.

Table 2 shows the number of years of participation, the number of EQAs, the average number of participants in each EQA, and the total number of EQA results for each analyte.

Table 2. Overview of data. The table shows an overview of the data included in paper I and II. Years of participation, number of EQA schemes, average number of participants in each EQA scheme, total number of EQA results.

Paper	Analyte	Years	EQA schemes (n)	Participants (n)	EQA results (n)
I	U-albumin	15	15	1159	69540
II	CRP	10	19	2134	162184
II	Glucose	10	19	2357	179132
II	Hb	10	19	2271	172596

In paper III, no EQA participants were included.

3.2 Measurement procedures

In paper I, the semi-quantitative POC Clinitek MPs (four MPs) and the Micral-Test (one MP) were included along with nine quantitative POC MPs (Table 3).

In paper II, 11 CRP, 41 glucose, and 18 Hb quantitative POC MPs were entered into the study.

In paper III, the most frequently used POC MPs for CRP, glucose, and Hb were included in the study (i.e., five CRP, seven glucose, and five Hb MPs). Furthermore, the accredited hospital MPs Cobas 6000 (Roche Diagnostics, Switzerland) (CRP, glucose) and Advia 2120 (Siemens Healthcare GmbH, Germany) (Hb), both located at the laboratory at Haraldsplass Deaconess Hospital (Bergen, Norway), were entered into the study. The ERM/DA474/IFCC CRM (CRP) and the SRM 965b (glucose) were used to validate the trueness of Cobas 6000 in the commutability study. The Advia 2120 (Hb) has good documented analytical quality through EQAs.

Table 3. Overview of the POC MPs. Table 3 shows an overview of the included MPs and manufactures in the three studies. (Next page.)

	Manufacturer	Measurement procedure	Measurement procedure version		Paper
Urine-albumin	**Siemens Healthcare Diagnostics**	Clinitek	Status		I
			Status+		
			50		
			100		
		DCA	2000 Analyzer		
			2000+ Analyzer		
			Vantage Analyzer		
	Roche Diagnostics GmbH	MICRAL-TEST			
	Axis-Shield PoC AS	Afinion AS 100 Analyzer			
		NycoCard U-Albumin			
	HemoCue AB	HemoCue	Urine Albumin		
			Albumin 201		
	Orion Diagnostica	QuikRead	U-ALB		
CRP	**Axis-Shield PoC AS**	Afinion AS 100 Analyzer			II, III
		NycoCard	CRP Single Test		II
				Reader II	
	Orion Diagnostica	QuikRead			
		QuikRead	101		
				CRP 1 min	
			GO		II, III
				CRP+Hb	
	HORIBA Medical	ABX Micros	CRP		II
			CRP 200		II, III
	Boditech Med Incorporated	i-CHROMA Reader			
Glucose	**Roche Diagnostics GmbH**	Accu-Chek	Aviva		II
				Nano	
			Performa		II, III
		Accutrend	Sensor		II
			Sensor		
	HemoCue AB	HemoCue	Glucose 201		II
				Plasma calibrated	
				Whole-blood calibrated	
				DM RT	
				RT	
			Glucose 201+		II, III
				Plasma calibrated	III
				Whole-blood calibrated	
			Glucose Analyzer	Plasma calibrated	II
				Whole-blood calibrated	
			Monitor		II, III
	Bayer Healthcare	Ascensia	Contour		II
				with Microfil REF 7085A	
				with Microfil REF 7085	
			Elite		II, III
				XL	
		Contour	with Ascensia Microfil REF 7085A		II
			Next		II, III
			XT		
		Glucometer Elite			II
	Abbott Diabetes Care Inc	Freestyle	Freedom Lite		II, III
			Lite		II
		Medisense	Precision Xtra		
				with Xtra plus	
				Xtra with prec H	
			Precision Xceed		
			Xceed		
				with Xtra plus	
				with prec H	
	LifeScan Inc	Gluco Touch			
		One Touch	Ultra		
				Easy	
				Smart	
				2	
			Vita		
Hb	**Siemens Healthcare Diagnostics**	Advia 60			II
	HemoCue AB	HemoCue	B-hemoglobin		
			Hb 201		
			Hb 201+		II, III
			Hb 201 DM		II
	Orion Diagnostica	QuikRead	GO		
				CRP+Hb	II, III
	HORIBA Medical	ABX	Micros	CRP	II
				CRP 200	II, III
			Microsemi		II
	Boule Medical	Swelab Alfa			
		AutoCounter AC920			
	EKF Diagnostic GmbH	HemoControl			II, III
	Sysmex Corporation	Sysmex	PocH-100i		II
			K1000		
			KX-21		
			KX-21N		
	Biotest Medizintechnik GmbH	Biotest Hemoglobin Tester			

3.2.1 EQA material

The EQA material used in paper I was human urine acquired from volunteers with normal or increased excretion of u-albumin. The EQA material was manufactured at Noklus before being stored at -80°C for a period of up to two years. Normal urine was added to the u-albumin urine before sterile filtration and separation into aliquots to achieve the desired concentrations. From 1998–2012 the u-albumin concentration in the EQA material was between 15 and 130 mg/L.

The EQA material in papers II and III was AB0-compatible EDTA venous whole blood collected from healthy blood donors at the blood bank and stored in plastic bags (Haukeland University Hospital). At Noklus (Haraldsplass Deaconess Hospital), the whole blood was pooled, and plasma or erythrocytes were added to attain the desired Hb concentrations. Human CRP (in.vent Diagnostica) was added to obtain suitable concentrations of CRP, and D(+)-glucose monohydrate dissolved in sodium chloride (VWR) was added to achieve the intended glucose concentrations. Iodoacetic acid sodium salt (VWR) and chloramphenicol succinate (VWR) were used to stabilize the glucose concentration and prevent bacterial growth. In paper III, the whole blood EQA material was kept refrigerated overnight before being divided into 2 mL cryovials (Sarstedt) and stored at room temperature until the next day to resemble typical mail deliveries to the participants.

In 2006, Noklus started producing its own in-house EQA material, which was subsequently distributed to all the participants analyzing POC CRP and Hb, and 53% of the participants analyzing POC glucose. The remaining laboratories that performed POC glucose received the commercial EQA material, Sugar-Chex Proficiency (Streck Laboratories). From 2006 to 2015, the concentration of the EQA material varied between 8–92 mg/L for CRP, 3–23 mmol/L for glucose, and 9–16 g/dL for Hb. For the commutability study in paper III, two concentrations of Hb and three concentrations of glucose and CRP were produced to cover the clinically relevant ranges. The following concentrations of the EQA material were produced for the study 23, 58, and 73 mg/L for CRP 7.0, 13.0, and 17.0 mmol/L for glucose and 9.6 and 13.2 g/dL for Hb.

The EQA material's homogeneity and stability were documented in accordance with ISO 13528 (78).

3.2.2 EQA surveys

In the EQA schemes in papers I and II, each of the participating laboratories obtained two EQA samples at different concentrations by mail. The participants were asked to store the EQA materials in a refrigerator and then allow the samples to reach room temperature before analyzing each EQA sample on two separate days. The participants analyzed each EQA sample twice using quantitative CRP, glucose, Hb, or u-albumin MPs, while participants using semi-quantitative u-albumin MPs analyzed each sample once. Within six days, they reported their analytical results to Noklus, along with information about the POC MP used, the lot number, the expiration dates of the reagents, the profession of the person who performed the analysis, how often IQC was performed, and finally, the number of patient samples performed at the laboratory each week.

3.2.3 Target values

In the EQA schemes in papers I and II, method-specific target values, calculated as medians after excluding outliers more than ± 3 SDs from the median, were used to assess the participants' performance for POC CRP, glucose, and Hb, in addition to the quantitative u-albumin MPs. For each of the two EQA material concentrations, the mean of duplicate measurements, calculated from the participants' EQA results, was assessed as "very good" when the mean concentration of u-albumin was within 7% of the target interval (target value ± 2 mg/L), and "poor" for results exceeding 15% of the target interval. The CRP participants' results were assessed as "good" when the mean concentration of CRP was within the target interval (target value ± 2 mg/L) ± 8% and "poor" for results exceeding the target interval ± 15%. Glucose participants were evaluated as "good" when the mean concentration of glucose was within 5% of the target interval (target value ± 2 mg/dL [±0.1 mmol/L]), and "poor" for results

exceeding 10% of the target interval. Hb participants were evaluated as “good” when the mean concentration of Hb was within the target interval (target value ± 0.1 g/dL) ± 3% and “poor” for results exceeding the target interval ± 8%. The participant performance was evaluated as “acceptable” for results between the limits for “poor” and “good” for all three analytes.

Participants using semiquantitative MPs reported their results in one of four possible categories. The category containing >30% of the semi-quantitative answers was used as the target value and represented the “good” category, the next category was labelled “questionable” and the remaining two categories were labelled “poor”.

In the statistical analyses in papers I and II, participants were characterized as “good” when they acquired assessments of “good” on both EQA samples, “acceptable” when they received an assessment of “acceptable” on one EQA sample and an “acceptable” or “good” assessment on the other, and “poor” for all other result combinations.

3.2.4 Patient samples and measurements

In paper III, the procedure for evaluating the commutability of a whole blood EQA material, delineated in the CLSI Guideline EP14-A3, was utilized (108). In agreement with the guideline document, the patient samples were gathered to cover the concentration range of the EQA material (Haraldsplass Deaconess Hospital). Altogether, 22 CRP, 23 glucose, and 25 Hb patient samples were collected. The CRP and Hb samples were leftovers from routine specimens at the laboratory, while the glucose samples were fresh capillary whole blood samples collected from healthy volunteers or volunteers diagnosed with diabetes mellitus. The glucose specimens were sampled by a biomedical laboratory scientist in a predetermined order to attain patient samples for both the POC MPs and the hospital MPs, the sample for the hospital MP, followed by samples for the POC MP, and then the sample for the hospital MP. Before and after each measurement sequence, IQC was performed on all POC MPs.

In compliance with EP14-A3, triplicate measurements of the patient samples with the EQA material randomly interspersed were performed on both the MPs for the POC and the hospital. EDTA blood was used for all the measurements performed on POC MPs, in addition to the analysis of Hb for the hospital MP. Serum was used for the analysis of CRP and glucose for the hospital MP. The patient samples were measured directly after collection, while the EQA material was analyzed two days after production, which included overnight storage at room temperature to imitate the mail delivery of the EQA samples to the participating laboratories. All measurements were performed at Noklus (POC MPs) and at the laboratory of Haraldsplass Deaconess Hospital (hospital MPs).

3.2.5 Statistics

In papers I and II, cross-tabulations and odds ratios (ORs) were calculated to evaluate the impact of performing the EQA. The EQA results categorized as "good" were compared to the results categorized as "poor" or "acceptable", and a two-sided chi-squared test and a Fisher's exact test were used for the computations. The level of statistical significance was set at $p < 0.05$.

In paper I, logistic regression was applied to each EQA survey. The participants' trueness and precision assessments were dependent variables, and the independent variables were the MP used, frequency of IQC, the profession of the employer performing the analysis, the number of GPs working at the GP office (until 2006), and the number of analyses performed at the laboratory each month (from 2009). The cutoff for statistical significance was set at $p < 0.05$.

In paper II, a binomial logistic regression was carried out after transforming the participants' assessments into a binary variable to compare good performance to poor/acceptable performance. The independent variables covered in the analysis were the number of participants in an EQA scheme, the type of participant, the profession of the employee analyzing, the number of analyses conducted each week, the frequency of performing IQC and when IQCs were performed, whether the MP was changed, and

the expiration dates on the reagents used. The cutoff for statistical significance was set at $p < 0.05$.

In paper III, the statistical analyses of the commutability of the EQA materials were conducted in accordance with the EP14-A3 guideline. First, outliers within the triplicate measurements were excluded from further analysis, then the homogeneity of variance was evaluated and, if applicable, the results were $\log_{10}$-transformed. The linearity between the paired MPs was assessed visually before difference plots were displayed for every MP combination, and a $\log_{10}$ transformation was utilized when required. Finally, outliers between the MPs were excluded. Deming regression analysis based on the mean of triplicate patient sample measurements ($\log_{10}$-transformed or untransformed values), was performed for each analyte and each MP combination. A 95% PI was calculated and plotted graphically, together with the mean values from the triplicate measurements of the EQA samples. The EQA material was considered commutable if the means of each of the EQA samples were within or touching the PI limits, otherwise, it was regarded as noncommutable.

The calculations were performed using Excel (version 14.4.8), SPSS (version 19.0, and 22.0), and the following packages of the software R (version 3.1.2, R Development Core Team, 2007): mcr (method comparison regression, version 1.2.1), methComp (functions for analyzing the agreement in method comparison studies, version 1.22.2), and boot (bootstrap functions, version 1.3.13).

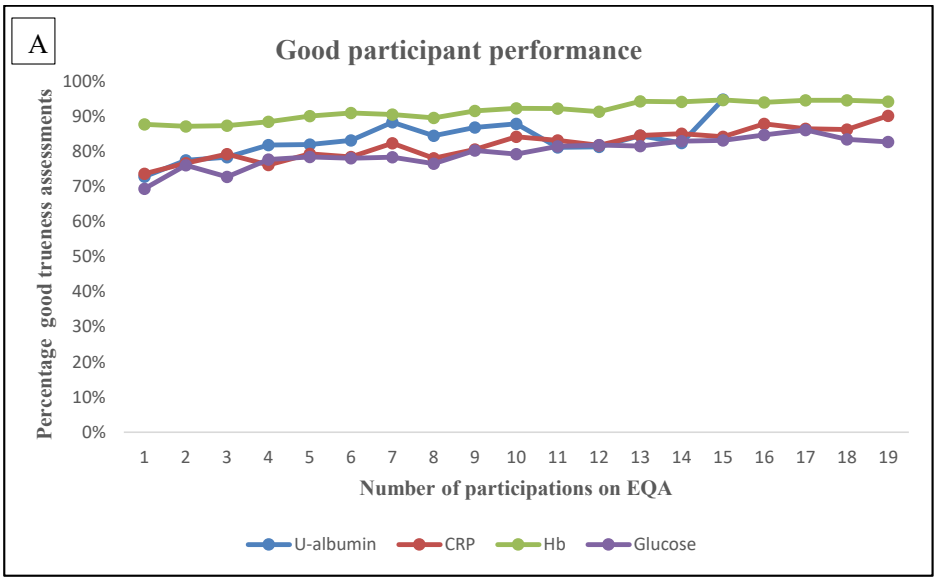
4. Results

4.1 The quality improvement system's effect on participant performance

In papers I and II, the percentage of participants who obtained a rating of good performance increased, and there was a similar decrease for those with poor performance, for all analytes assessed. The length of time the participants had been a part of the Noklus quality improvement system seemed to have an influence on their performance.

In paper I, the percentage of GP offices receiving an assessment of good or poor trueness increased and decreased, respectively, and seemed to be influenced by the length of time they had participated in EQA (Figure 9A and Figure 9B). When comparing old to new participants, the OR for receiving the assessment of good trueness compared to poor trueness generally increased the longer the GP office had participated in the EQA (i.e., comparison of participation for the third and ninth times yielded ORs of 1.5 [95% CI: 1.1–2.1] and 4.2 [95% CI: 2.2–7.8], respectively). Significantly more old participants, unlike new participants, were assessed as good compared to poor for precision (OR: 1.5 (95% CI: 1.2–1.8)).

In paper II, the percentage of participants with an assessment of good trueness on the CRP analysis increased from 73% to 90%, while the percentage of participants with poor trueness assessments decreased from 12% to 3%. For glucose, good trueness assessments increased from 69% to 83%, and poor trueness assessments decreased from 10% to 3%. For Hb, the number of good trueness assessments increased from 88% to 94%, and the number of poor trueness assessments decreased from 3% to 0% after performing EQA for the first and nineteenth times, respectively (Figure 9A and Figure 9B). The overall OR for receiving good performance versus poor performance for participants who had taken part in a minimum of two earlier EQAs, compared to participants who were taking part for the first time, were 2.12 (95% CI: 1.66–2.70) for CRP, 1.62 (95% CI: 1.22–2.15) for glucose, and 1.84 (95% CI: 1.27–2.68) for Hb.



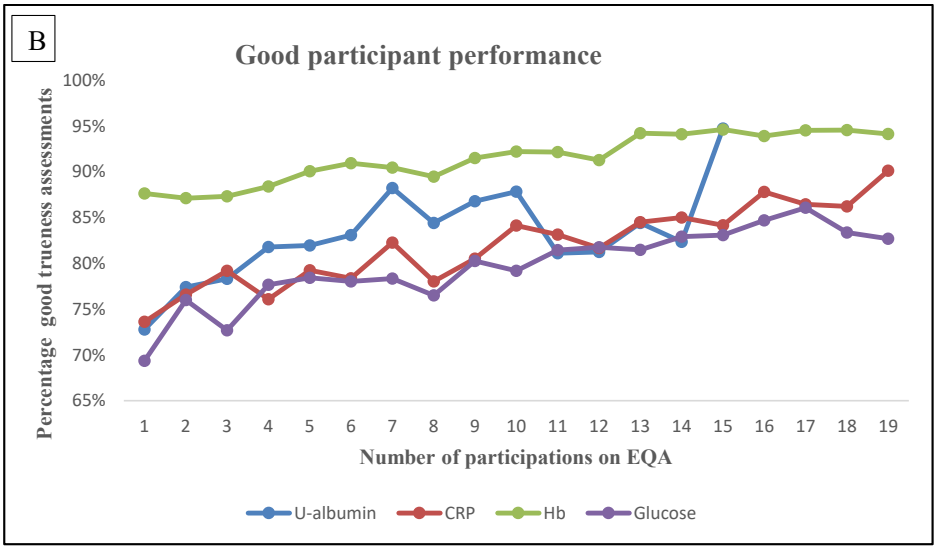


Figure 9. Participation over time. The figures show good participant performance in relation to number of years participating in EQA. The blue line shows u-albumin (15 EQAs), the red line shows CRP (19 EQAs), the green line shows Hb (19 EQAs), and the purple line shows glucose (19 EQAs). In Figure 9A the y-axis ranges from 0-100% whilst in Figure 9B the y-axis ranges from 65-100%. The figures illustrate a gradual increase of participants obtaining good participant performance.

4.2 Factors associated with good participant performance

In paper I, it was concluded that participants using quantitative MPs who obtained a poor trueness assessment received significantly fewer poor trueness assessments the year after changing their MP, regardless of whether they chose another edition of the same MP or a new type of quantitative MP. The change in MP resulted in significantly fewer good trueness assessments for participants who initially obtained an assessment of good trueness. Participants who used quantitative MPs involving reagents with over three months left before their expiration date, performed significantly better than those using reagents expiring within three months ($p < 0.001$) or reagents that already had expired ($p < 0.001$).

Several practice characteristics changed significantly between 1998 and 2012. The percentage of participants who never performed IQC decreased from 48% to 6%. The percentage of participants choosing a quantitative MP (DCA) increased, and the percentage choosing a semi-quantitative MP (MICRAL–TEST) decreased. The DCA MPs performed better than each of the other quantitative MPs for both trueness and precision. The profession of the operator, the frequency of IQC, the number of GPs at the GP office, and the number of analyses performed each month did not significantly influence the participants' assessments.

In paper II, a logistic regression analysis was implemented with a binary variable for good versus poor/acceptable performance. The study identified a number of independent factors that were associated with good participant performance, performing EQA more than once, the operator was a medical laboratory scientist, 10 or more analyses were performed each week, there was weekly IQC, and the kind of MP that was used. For CRP and Hb, the GP offices performed significantly better than the other participants, whereas nursing homes performed best in the EQA regarding glucose. The participants using reagents with more than 60 days until their expiration date performed significantly better than participants using reagents with 60 or fewer days until the expiration date of their reagents in the EQAs for glucose and Hb. Whether the MP was changed did not influence performance.

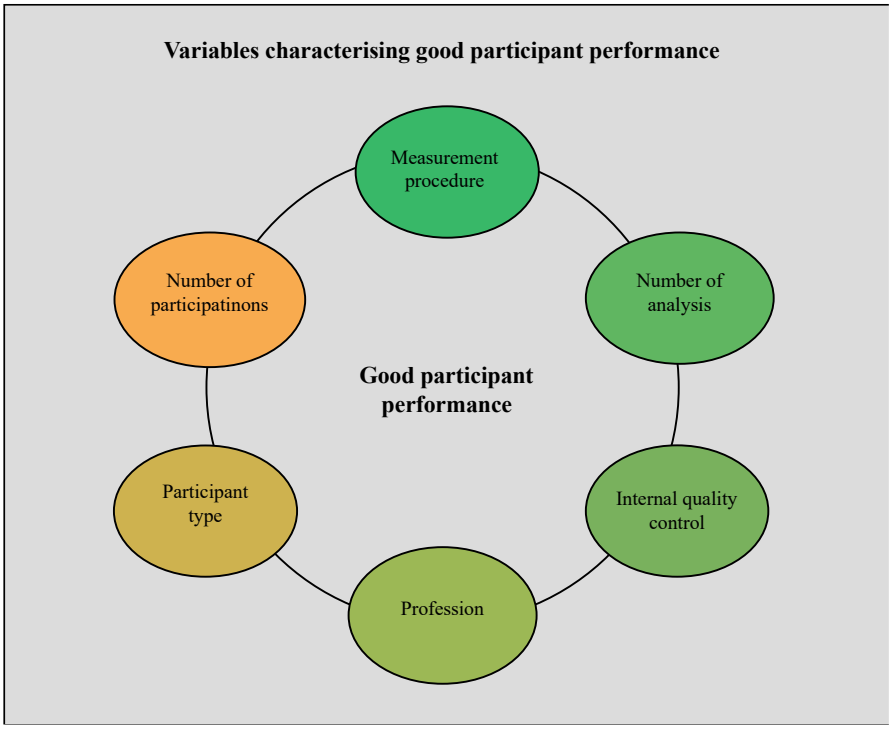


Figure 10. The figure illustrates which variables characterized good participant performance.

4.3 Commutability of the EQA material

When the CRP and Hb POC MPs were compared with the hospital MPs, the EQA material was commutable for all combinations except for the QuikRead GO at the highest concentration of CRP (Table 4).

For glucose, the EQA material was commutable between the hospital MPs and all POC MPs at the lowest concentration. For the two higher concentrations, commutability was only demonstrated for the Contour and HemoCue MPs (Table 4).

Pairwise comparisons of the POC MPs revealed that the EQA material was commutable for nearly all combinations for CRP and Hb. For glucose, 37 of 63 combinations showed commutability, mainly at the lowest concentration.

Table 4. Commutability of the EQA material. The table shows for which MP combinations the EQA materials was commutable (green) and noncommutable (blue). Only POC MP in combination with the hospital MPs is shown.

Commutable			
Noncommutable			
Hemoglobin	**9.6 g/dL**	**13.2 g/dL**	
Measurement procedure	**Advia 2120**	**Advia 2120**	
ABX Micros CRP 200			
HemoControl			
HemoCue Hb 201+			
Sysmex PocH-100i			
QuikRead GO CRP + Hb			
CRP	**23 mg/L**	**58 mg/L**	**73 mg/L**
Measurement procedure	**Cobas 6000**	**Cobas 6000**	**Cobas 6000**
ABX Micros CRP 200			
Afinion AS 100 Analyzer			
i-Chroma			
NycoCard CRP Single Test			
QuikRead GO CRP + Hb			
Glucose	**7 mmol/L**	**13 mmol/L**	**17 mmol/L**
Measurement procedure	**Cobas 6000**	**Cobas 6000**	**Cobas 6000**
Accu-Chek Performa			
Ascensia Contour			
Contour			
Contour XT			
FreeStyle Freedom Lite			
HemoCue Glucose 201+			
HemoCue Glucose 201RT			

5. Discussion

The present thesis evaluated participants' assessments in different EQA schemes over time and assessed which factors are associated with good performance. Additionally, the commutability of the whole blood EQA control material used for the EQA scheme was examined.

5.1 Methodological considerations

5.1.1 The observational descriptive studies

Study population

In paper I, GP offices performing u-albumin analysis were included in the study. The EQA scheme for u-albumin at Noklus, however, includes analysis of both, participants' u-albumin and u-albumin/creatinine ratio. An important aim of the study was to make use of Noklus' data to assess the variables associated with good performance. For the u- albumin analysis, 14 different MPs took part throughout the entire study period compared to only one MP analyzing the u-albumin/creatinine ratio in the same time interval. The type of MP used by the participants was expected to be an important variable affecting performance, and hence, u-albumin data were chosen for further analysis in study.

In paper II, all participants performing either CRP, glucose, or Hb measurement in primary health care in Norway were included, GP offices, nursing homes, emergency primary-healthcare facilities, family healthcare centers, and providers of home care and occupational healthcare. Different analytes and participant types were included, which increased the overall data set and allowed the evaluation of the potential differences in performance between analytes and types of participants. The study period included data from 2006 to 2015. In 2006, Noklus started producing their own in-house whole blood EQA material for the joint CRP, glucose, and Hb EQA scheme. Limiting the participant results to the same EQA material reduced the heterogeneity of the study group and increased the reliability of the study design as the EQA material is an essential component of the scheme and studies. If the EQA material were to differ among the

study groups, it would be difficult to recognize real differences in participant performance, which, however, can be related to the importance of commutability of the EQA material when assessing differences in the performances of MPs.

To be included in the studies in paper I and II, the participants' MPs had to be used by at least 8 participants. This criterion is routinely used in the EQAs at Noklus when a method-specific target value is calculated to reduce its uncertainty. In the studies, the overall participant performance was used as the dependent variable and to ensure reliability only participants reporting numeric results on both replicate measurements at both levels (four analytical results) in the EQA scheme was included. This inclusion criterion provided a more reliable estimate of the participants' performance and made it possible to obtain an estimate of the MPs precision. The number of participants excluded due to these selection criteria was, however, negligible, and therefore, practically all participants in the Noklus EQA schemes for u-albumin, CRP, glucose, and Hb were included in the respective studies. The amount of data (nearly 600 000 analytical results in total), along with the knowledge that 99% GP offices and 96% nursing homes in primary care are enrolled in Noklus, resulted in a large cohort of participants included in the respective studies and an increased likelihood for the data being representative of primary-care laboratories in Norway. In addition, there should be no selection or information bias in the dataset because nearly all participants were included, providing the same set of information. Based on the population included, the external validity of the studies was considered to be strong.

Participating in EQA scheme over time

There have been, as abovementioned, concerns about the management of POC quality in primary care (111-115). Its importance, however, has also been recognized, and in paper I and II in this thesis, the effect of quality management of POC in primary health care has been evaluated. The performance in EQA over time was assessed using an observational study design. The retrospective studies were conducted using data from already performed EQAs for u-albumin, CRP, glucose, and Hb. The participants'

overall performances on the different EQAs were linked to the unique laboratory number, making it possible to observe the effect of participation in EQAs over several years. A recent review by Price et al., in which whether quality management of POC testing leads to better test results was evaluated, identified five studies, all observational in nature, in which the performance of POC testing in EQAs was used to evaluate the improvement over time (8).

The study designs for paper I and II were chosen to benefit from the extensive set of data Noklus acquires on EQA schemes in primary care in Norway. Noklus is, in fact, one of few organizations worldwide in possession of EQA data covering such a long period of time. Characteristics of the dataset favoring comparability throughout the entire study period and the reliability of data were the similarities in the obtained information from the participants, and the unaltered APS made the participants' assessments comparable from one year to another. The study design enabled us to follow the participants retrospectively over a relatively long period of time and allowed us to find long-term patterns in their performance.

In paper I, the performance in EQA over time was assessed by comparing old and new participants using the OR. Factors associated with good analytical quality were changing MP when the initial performance was assessed as poor and using reagents that were not expired or had more than three months left to the expiration date. Observational studies, however, run the risk of confounding bias, which might affect the outcome of the study. Confounding was addressed further in paper II, in which logistic regression analysis was performed to assess whether the variables were independent predictors for good participant performance. The regression identified the number of participations in the EQA, performing IQC weekly, performing more than 10 tests weekly, using recommended MPs, and having laboratory qualified operators as independent predictors for good participant performance. Although these variables are important aspects registered in the EQAs by Noklus, they are not necessarily the only variables contributing to good performance. As opposed to experimental studies, in retrospective observational studies, you have a lower ability to control for confounders. Factors other than those included in the estimation model could

potentially influence the performance, e.g., participation in courses or training of the operators, which, due to lack of data, was not included in the analysis (116). Participation in courses or training is one example of threats to the internal validity, where something else besides the participation in the EQA scheme accounts for some of the observed improvement. Such factors could be specific events, as in the case of participation in training, or they could be part of a process of growth and spread of information and knowledge over time. Because the research was conducted in a real-world setting, the data collected would, naturally, be affected by an increasing number of old participants compared to new participants the longer the study lasted, and the old participants would also, in general, have their performance recorded later in time than the new participants. If there is a time trend of continuous improvement that is unrelated to participation in EQA schemes, in the dataset, the participant performance would be expected to improve with an unknown magnitude independent of the EQA schemes distributed. The time variable was, however, assessed by evaluating whether the number of good participant performances gradually increased with each EQA distributed. The percentage of good performances showed, on contrary, an irregular curve with no obvious time trends. Furthermore, all data available in the EQA database were considered to increase the internal validity of the study.

In paper I and II, the same participants' continuous measurements in primary care were studied over 15 and 9 years, respectively. A prerequisite for the statistical tests used was that the observations were independent because ordinary regression models do not consider any possible dependency between repeated measurements. An alternative to the logistic regression model used in study II was to use other statistical methods used in longitudinally observational studies like generalized estimated equations (GEE) or the mixed effects model. These models take into account an unknown correlation between the measurements. In addition, cluster analysis could have been performed to evaluate whether some groups of objects in a study are more similar than objects in other groups. However, since the participants were GP offices, nursing homes etc., and not the same individuals, the data were, after discussions, presumed to be independent of each other. Additionally, Noklus emphasizes in the EQAs that the EQA material

should be handled and analyzed as routine samples, and measurements are to be performed by different employees normally performing the POC testing.

An alternative approach to the descriptive study design would have been to include a comparison group and performing an experimental study to evaluate the data using two randomized cohorts. Then, the effect of participation in EQAs, different independent variables, the time trend, and potential confounders could have been more closely scrutinized. Such a design would, however, be impractical as 99% of all GP offices and 96% of all nursing homes in Norway at the time of the studies already took part in the EQA voluntarily, and a control group would consequently be difficult to establish. An alternative could be to let some of the EQA participants take part in the EQA scheme but not receive feedback on their results, nor been followed up in cases of poor assessments. An evaluation of the control groups' assessments compared to the ones receiving feedback and support could then be performed. Such a study design would, however, be impractical as Noklus' data could not be used to its full potential because the participants would have to be followed up in a prospective study. Additionally, the study design would be unethical. Provided the EQA schemes are performed to ensure laboratory quality and patient safety, knowledge regarding poor POC test results in a facility without intervening in the name of research could potentially lead to patient harm, which is the opposite of what the current research intended to do. Finally, to be able to assess differences between new and old participants, the study design, in our case, had to be retrospective. Because the larger share of primary care today takes part in EQAs, results from new participants could only be obtained by gathering data from when EQA was first introduced in Norway. In addition, to follow the participants over such a long period of time and assessing whether there is an improvement in the POC quality over time, not only whether there is an effect, a longitudinal observation study is the preferred study design.

5.1.2 **The CLSI EP-14-A3 standard for evaluating commutability**

The commutability of the whole blood EQA material was examined utilizing the guidelines designated in CLSI EP-14-A3 (108).

Measurement procedures

In paper III, the most commonly used POC MPs for CRP, glucose, and Hb in primary care in Norway were included in the commutability study. Ideally, all MPs used in an EQA scheme should be covered for in such a study to verify whether the EQA material is commutable for all POC MPs for which the EQA material is ought to be used (55). Furthermore, the CLSI standard recommends that every new batch of EQA material produced is tested for commutability. Performing a commutability study is, nevertheless, both time consuming and costly, especially in cases where whole blood EQA material is distributed to the participants. The batch of whole blood EQA material then needs to be freshly prepared prior to each EQA scheme. If the recommendations of testing every new batch and every new MP were followed, several commutability studies had to be performed yearly. A reasonable approach, which has also been used by Noklus, has been to test the most commonly used MPs in the EQA scheme and assume commutability of new batches of the EQA material, provided they are prepared in a similar manner (55). To cover for the most commonly used POC MPs in primary care for the analytes, five MPs for CRP and Hb and seven MPs for glucose were included in the commutability study. The proportions of participants using one of these MPs in the EQA scheme at the time of the commutability study were 98%, 87%, and 97% for CRP, glucose, and Hb, respectively. The external validity of the commutability study was therefore regarded to be strong.

When assessing the commutability of the POC MPs, it can be an advantage to have a comparative MP for evaluation. Ideally, the comparative MP should have minimal matrix effects, and a reference MP is therefore preferable (117). For CRP, there is currently no reference MP, and the hospital MP, Cobas 6000, was therefore included in the commutability study along with the CRM ERM/DA474/IFCC CRM (94). The

glucose and Hb results are, on the other hand, traceable to the reference MPs ID GC-MS and HiCN, respectively (61, 101). The two reference MPs could, in theory, have been included in the commutability study. For glucose, however, method-specific target values are still used in the EQAs. The systematic bias is, however, calculated for MPs for which the EQA material has shown commutability and, in this process, Cobas along with the certified reference material NIST SRM 9965b is used. For Hb, the target value for the EQA material is established using the HiCN reference MP in Odense university Hospital (Odense, Denmark). For practical reasons, however, the frequently used hospital MP Advia 2120 was chosen as the designated comparison MP. Because the patient samples were to be analyzed freshly on both the POC and the designated comparison MPs, it was not possible to include the reference MP in the commutability study for comparison. The CLSI guidelines also argue for using a validated hospital MP in cases where a reference MP is not practical (117). Nevertheless, Advia has shown good analytical quality when assessed in EQA schemes, where a true value has been used. Consequently, there is no reason to believe the results from the commutability study are not valid.

Preanalytical challenges

The patient samples needed for CRP and Hb measurements could be collected from leftover routine samples, and after collection, they were freshly analyzed as a single batch within one day, which is recommended by the CLSI standard. For the analysis of glucose in fresh capillary blood samples, however, some adjustments had to be made to overcome the preanalytical challenges mainly associated with stability (118). The patient samples for the glucose study were, therefore, collected and analyzed over six days, which may have influenced the day-to-day variation. Several efforts were, however, undertaken to minimize the variation and help in maintaining the internal validity of the study. An experienced biomedical laboratory scientist performed sample collection by following current guidelines and analyzed the samples in the exact same order in every measurement sequence (119). The punctation site was heated prior to punctation to avoid squeezing which, as abovementioned, is one of the most common

confounding factors in capillary blood sampling. Furthermore, the first blood drop was discharged. The glucose samples that were to be analyzed on the hospital MP were immediately centrifuged and frozen to avoid glycolysis and a decrease in the glucose concentration. There have, however, been concerns that freezing might change the clinical samples' matrix (67, 120). Freezing could then have altered the matrix of the hospital samples, making the samples not comparable to the samples analyzed directly on the POC MPs. To limit the differences and maintain the validity of the study, plasma was frozen at -80°C for a short time and the procedure was performed in accordance with the National Institute Standards and Technology's protocol for storing CRMs (121). The plasma samples were also, as recommended by CLSI, analyzed on a single day to prevent any day-to-day variation influencing the data, which increased the validity of the study (108).

IQC was performed on all POC MPs before and after each measurement sequence. Finally, all cuvettes had the same lot number to prevent lot-to-lot variation because studies have indicated that noncommutability may occur between lots (122, 123). It is, nevertheless, important to emphasize that the recommendations from CLSI do not address the analyte per se but are general recommendations on how to perform a commutability study.

Statistical power

The CLSI 14-A3 recommends obtaining at least 20 patient samples for the commutability study, in the current study, 22 CRP, 23 glucose, and 25 Hb samples were collected. The patient samples and the number of replicates is important for the statistical power, i.e., the probability for determining whether the EQA material is commutable or not, and thus, for the internal validity of the study. The number of samples and replicates have been increased in studies to evaluate the effect on PI (124). The results, however, did not show narrowing of the PI, but an increased confidence in the PI obtained (124). Other methods like virtual pooling or calculations of the statistical power to determine the sufficient samples size have also been suggested to

improve the statistical power (124, 125). Although an increased number of samples or replicates could have increased the power of the study, 20 to 25 patient samples were the number of clinical samples that was practically achievable considering the recourses of the study and the number was sufficient following the advice of the CLSI guidelines (108).

Assessing commutability

According to the CLSI standard, a visual assessment of the variance of homogeneity, the linearity between MPs, and the degree to which the SDs increased with increasing concentration should be performed (108). These assessments are, as a result, to a certain extent, subjective. This is also the case when evaluating the Deming regression plots as the EP14-A3 standard does not give any guidance on how to interpret such data points or how many must fall within the PI's limits for the EQA material to be considered commutable overall. Difficulties arise particularly when the values of the EQA material fall on or close to the PI's limits. In such cases, the individual judgement can affect the outcome of the commutability study and thereby affect the study's validity. To increase the reliability of the study design, common rules for evaluating the data points were established. First, data points located on the prediction lines were assessed as commutable, whereas data points right outside, not touching the prediction lines, were assessed as noncommutable. Second, each of the EQA material's data points in one Deming regression plot had to be assessed as commutable for the EQA material to overall be assessed as commutable for the particular MP couple. In addition, all MP couples were discussed in plenary to ensure reliability. Our assessments on commutability in the current study, however, seemed to be in concordance with other studies on the evaluation of commutability using the CLSI guideline (126).

Noncommutability for the EQA material was primarily observed at the high concentration of glucose. Because Deming regression is a linear model, the regression line and, therefore, the PI, is determined by all the data points in the selected population. As the patient samples, in general, were dominated by samples with low

concentration of glucose, the regression line and, consequently, the PI, could be more appropriate for assessing EQA materials at a lower concentration range. Similar findings were also observed in a study by Liu et al. (127). To increase the validity of the study, one should aim to have approximately the same number of patient samples at all concentration levels. However, in practice, this may be challenging when the patient samples are to be collected fresh and spiking must be avoided.

In 2018, the IFCC working group on commutability launched the IFCC difference in bias approach for assessing commutability of reference and EQA materials (86, 109, 110). In contrast to the CLSI method that utilized the statistical distribution of patient results when comparing two MPs to each other, the new recommendations suggested evaluating the disagreement in bias between the patient samples and the EQA material and, additionally, calculating error bars for uncertainty (18). Using error bars could have been an advantage in our study as the assessment of data points located on or near the PI limits could have been easier. A standardized way of assessing such data points would have increased the validity of the study. When using the CLSI method, calculating the limits of the PIs depended on the width of the scatter of the clinical samples and, thereby, the random error of the measurements from each MP. The criteria for whether the EQA material was assessed as commutable would, therefore, differ with the MPs evaluated. The IFCC difference in bias method recommends applying fixed limits based on medical requirements. The method also suggests only including MPs documenting the acceptable precision and selectivity since a large imprecision would result in wide error bars, making the commutability assessment inconclusive. Though it is a key limitation of the CLSI method that it is more stringent for precise MPs compared to less precise MPs, excluding MPs can be problematic. It is therefore possible that different criteria should be used for evaluating commutability of CRM materials and EQA materials. If a MP holds a significant market share and is widely used by the EQA participants, excluding the MP from the study would make the results less valuable both for the EQA provider and the omitted participants and would decrease the internal validity of the study (128). In addition, focusing on medically relevant differences can be of great importance when assessing the agreement between

MPs or evaluating the commutability of reference materials or calibrators, however, it may not be of equal importance when evaluating the commutability of EQA materials. For EQA materials, in our opinion, the key focus should be on whether the EQA material behaves similarly or different to native patient samples routinely analyzed by the participants.

When medical requirements are used as criteria, they should ideally be based on outcome studies, c.f., the Milan criteria (66). Due to limited access to outcome studies, the APS used in laboratories are often based on biological variation, which are not based on medically relevant differences. In addition, using APS based entirely on biological variation may, for some analytes, be too strict, especially when evaluating EQA materials. The APS should, therefore, as mentioned earlier, take unique variations for EQA materials into consideration. Two studies recently assessed the commutability by applying both the CLSI and the IFCC methods (127). Whereas the first study, assessing a whole blood HbA_{1c} CRM, concluded that the two methods gave overall consistent results, the second study, assessing serum sodium and potassium, obtained a large share of inconclusive results using the IFCC bias approach compared to the CLSI standard (127, 129). One of the main reasons underlined by the authors is the medical requirement criteria being too strict, which has also been argued by others (130, 131). Although none of the methods may be ideal for assessing the commutability of different materials, in our case, the IFCC method was not available at the time of the study and, therefore, only the CLSI standard was applied in the commutability study.

5.2 Main results and implications

POC testing has, in recent years, become increasingly available and especially with the current Covid-19 pandemic, rapid tests have been a key to quick tracing, isolating contacts, and helping break the transmission chain. Governments around the globe are spending billions on testing and tracing for Covid-19. One example is the UK government, which so far has spent over £10bn alone (132). To help stop outbreaks early and to ease lockdown restrictions, the UK is also offering free rapid tests for everyone across the country to be tested twice a week. For POC tests to be fit for

purpose, they must be both reliable and effective, not only in the laboratory setting of the manufacturer but in the range of different locations where the POC tests are performed. To recurrently ensure the quality of POC testing in the real-life setting, EQA schemes are used to assess the performances of participants and the MPs, and to enable this assessment, the quality of the EQA material is crucial (133, 134).

5.2.1 Participating in EQA over time

POC testing is typically performed outside the laboratories, near the patient. However, it is still defined as laboratory testing and should be treated as such. The entire testing loop must therefore be applied also to POC testing, including all aspects from the pre-preanalytical to the post-postanalytical phases.

The main focus of EQAs is to verify that the analytical quality conforms to the expectations for the quality needed for patient care (55). The scope of the EQA organizations has, however, also expanded to include both pre- and postanalytical components (55). Noklus comprises such a total quality system to cover the entire testing loop for POC in primary health care in Norway (75).

In paper I and II, data from Noklus were evaluated to assess whether performance in EQAs over time resulted in improved quality. The analytical quality of the participants' performances was used as the study outcome, and for all four analytes, the data showed a gradual increase in the number of participants who achieved good performance in the EQAs. Correspondingly, a decrease was observed for the number of participants whose performances were poor. The number of times a participant participated in EQA was also recognized as a key factor associated with good performance. Although EQA is recognized as an essential part of the laboratory's quality management, relatively few studies have addressed whether participation in EQAs over time can give rise to improved quality. A systematic review from 2018 identified five studies using valid outcome measures to evaluate whether performance in EQAs improved the quality of POC testing in primary care (8). Three of these studies were performed at Noklus, two of which were a part of this thesis. In all studies, relevant outcome measures were

established by creating summary performance criteria, which were based on the participant's EQA results assessed against predefined APS. In two of the studies, the imprecision of POC MPs was evaluated after participation in EQAs over 3, 5, and 9 years, respectively, to the performance of the central laboratory. Both studies showed a decrease in the MP imprecision throughout the study period and concluded that the POC performance after the study period was equal to the one at central laboratories (135, 136). In the third study of Sølvik *et.al* performed at Noklus, two POC MPs for HbA_{1c} were evaluated on whether they met specified target values for both bias and imprecision (137). The study found an improvement over time and concluded that the participants had about 90% probability of meeting the quality specifications when one MP was used (137).

Including trueness as a study outcome when evaluating performance in EQAs over time is a strength. Because the number of replicates in the EQAs often are limited, the trueness is often a more reliable measurement than the CV, especially in cases where there is a true target value, like for Hb in paper II. Another reason for the trueness to be a robust outcome measure is the fact that the target value, and thereby the trueness result, is unknown for the participants. If one of the two replicates show a very divergent result, it may be more likely that the participants reanalyze the samples due to lack of precision than due to lack of trueness. This may also be linked to the Hawthorne effect, which may occur because of people know they are being watched (138). However, participating in Noklus EQAs is educational and not required for accreditation. In addition, the EQA material is distributed to the participants frequently, and Noklus emphasizes in the schemes that the samples are to be handled like routine clinical samples.

Another strength of the studies in paper I and II is the fact that the same EQA material was used throughout the study period. Similar studies have mixed EQA results from different EQA schemes and different EQA materials, both liquid and lyophilized (136, 139). As indicated in these studies, processed EQA materials may have unknown changes in their matrix, which may cause them to behave differently on different POC MPs (67, 74). The issue concerning matrix effects and commutability of the EQA

materials at Noklus was addressed in paper III of this thesis. Because the EQA material, at the time of the study of papers I and II, did not have documented commutability, the observed improvement was related to an improvement in participant performance or the between participant variation for one MP over time. In the case the EQA material was known to be commutable, the improvement reported in the studies would also reflect an improvement in the MP performance over time. Because four different analytes were assessed in the studies, it was possible to evaluate whether the effect of participation applied to different analytes and to what extent the effect was applied. Interestingly, the effect was not so much pronounced for Hb, which had a less of an increase in good performance. The participant EQA results for Hb have, for several years and through the entire study period, been evaluated against a true value. This has probably resulted in standardization and harmonization of Hb results, and the effect of participating in EQA schemes may therefore be more pronounced for analytes that do not have calibration traceability to a reference MP.

One of the main benefits of using POC is reduced TAT, which in turn may lead to faster diagnosis and treatment and may thereby reduce referrals to secondary care (140, 141). In some cases, it seems that using POC testing can lead to better patient care compared to hospital laboratory testing, as exemplified by Tollånes et al. for HbA_{1c} (142). There are, however, several concerns identified toward introducing POC in the primary healthcare setting (112). Some of the issues, such as concerns with the accuracy of the POC tests and quality management, are usually addressed by performing IQC and participating in EQA schemes. In a total quality management system, like Noklus, the participants are offered help and guidance when needed. For the EQA organization to have knowledge of which factors are important for good quality is therefore essential and was addressed in this thesis.

The MPs used in POC testing range from small benchtop MPs, which in practice are small-scale laboratory MPs, to small handheld portable MPs (20). Benchtop MPs are commonly used in GP offices, and the analyses are performed during the daytime by a limited number of operators. In contrast, the handheld MPs are used in a wide variety of settings, at all hours, and by the operator at work. In the study in paper II, the GP

offices achieved better performance than the other participants on the analysis usually performed using benchtop analyzers, whereas the nursing homes showed the best performance when performing the glucose analysis, normally performed using handheld MPs. At the GP offices one is more likely to find laboratory-qualified people employed. Laboratory training may reduce preanalytical factors because of knowledge of how to prepare the patient and how to perform sample collection and analytical factors by performing IQC and calibration. Errors in the postanalytical phase are often linked to the organizational settings and may not be related to laboratory education to the same extent. The fact that the nursing homes achieved the best quality when performing the glucose POC testing may be explained by the setting. Because glucometers are mainly used for self-monitoring, the test is more likely to be performed by staff at a nursing home than at a GP office. Performing the analysis frequently, which was another key factor associated with good performance, may, therefore, explain why nursing homes perform better than other participants in this context.

Performing IQC on a weekly basis was a key factor associated with good performance. IQC, similar to EQA, is an essential component of quality management in laboratory testing. As IQC is essential for detecting, reducing, and correcting errors in the analytical phase in real time, it is not surprising that performing IQC frequently, is associated with good performance in EQA schemes. Finally, as expected, the MP used was of great importance for the quality of POC testing. The MPs differ both in analytical quality and user-friendliness, which are of great importance for the POC MPs as they are used under changing locations often by non-laboratory operators. Noklus, together with the Scandinavian evaluation of laboratory equipment for point of care testing (SKUP), evaluates and recommends POC MPs for primary care (143). The studies included in the present thesis can inform the participants of the importance of choosing the right MP and the importance of performing IQC frequently.

5.2.2 Commutability of the EQA control material

Usually, processed or lyophilized EQA materials are used in EQA schemes as they have several attributes relevant to EQA providers. Examples of such attributes are large number of analytes that can be pooled together, these analytes are cost-effective and have excellent stability, which makes the material suitable for storage for later EQA schemes (67). However, matrix effects may interfere in the processed samples, thereby making them noncommutable (144, 145). To provide the participants with EQA samples that resemble native patient samples as much as possible is the aim of several EQA organizers, including Noklus. Different studies have addressed the importance of commutability in recent years, and many organizations have started evaluating the commutability of their materials. However, the studies are mainly assessing the commutability of reference materials being used as common calibrators, or the materials evaluated are chiefly serum or processed materials (127, 129, 146-148).

This study is, by our knowledge, the first to show results of a non-hemolyzed whole blood EQA material which has been stable during the entire analytical period and also been commutable.

A classification of EQA schemes into categories, depending on how well they can assess performance of the participants and MPs, was suggested earlier by Miller (55). Three essential characteristics of the EQA scheme are crucial for categorization, whereas the commutability of the EQA material is the most important. Before the commutability studies were conducted, all EQA schemes at Noklus were classified into category 5. As the commutability of the EQA material had not been evaluated, the schemes were limited to a peer-group assessment. Because of the bias between MPs, it was not possible to assess as the peer-group target value might be influenced by matrix effects (3). The results from the commutability study showed commutability for all CRP and Hb MPs, and because the CRP measurement is traceable to a CRM, and the Hb analysis has traceability to the HiCN reference method, a true common target value can now be calculated for these analytes (94, 101). Thus, participants using POC MPs for which the EQA material is commutable can now be classified in EQA category 1. Because the commutable EQA material, in theory, should have the same results for all

MPs and lots, the observed differences reflect real differences between the MPs used. In category 1, the participants can get an evaluation of the trueness and precision of their EQA result and the MP used, the reproducibility can be evaluated, and the EQA result can be used for standardization and harmonization efforts to achieve uniformity among MPs (55). Determining the EQA material's commutability is a considerable advantage for the EQA organization, the participants, and the IVD industry as it provides for more adequate guidance as to how true analytical differences between the MPs can be detected. Distributing commutable EQA materials is also essential for the standardization and harmonization efforts currently taking place. The results from the EQA scheme will provide information regarding the true errors and differences between MPs, and the EQA organizations then have a unique position to conduct surveillance of the harmonization efforts and also be able to detect analytes in need of harmonization (149). Such a harmonization was, for example, seen for apolipoprotein A-I after the development of an international reference material and a standardization program which enabled traceability. The national EQA scheme could, following the harmonization, confirm a significant reduction in interlaboratory variation from 35% to 10% (146).

For glucose, the commutability study revealed a differentiated response on the whole blood EQA material. Similar observations regarding glucose MPs have also been seen in other studies (122). For about half of the MPs that showed commutability at all concentrations, a true value can be calculated. The participants using these glucose POC MPs could then get their EQA results assessed compared to a true value. One cannot, however, expect all participants to have detailed knowledge regarding the commutability of the EQA material and how it influences the calculation of the target value. To explain why different target values are used for the same analyte can, therefore, be challenging. As a result, a peer-group target value is still used in the Noklus EQA scheme for glucose. However, Noklus calculates the systematic deviation from the true target value for the POC MPs for which the EQA material showed commutability. The information is available for IVD manufacturers and participants if needed and is also used by Noklus to provide recommendations on which MPs to use e.g., in screening for gestational diabetes (76).

An additional finding in the commutability study of glucose was that all analytes showed commutability at the lowest concentration. A similar result was demonstrated in a commutability study of a CRM for HbA_{1c} by Liu et al. (127), in which in the two cases where the material was assessed noncommutable, it was at high concentrations. The finding emphasizes two crucial issues. The first is the importance of including more than one concentration when assessing a reference or EQA material. In the study in paper III, three CRP and glucose concentrations and two Hb concentrations were examined. The present study showed that inclusion of several EQA samples at different concentrations is of utmost importance as commutability at one concentration does not naturally denote commutability at other concentrations.

The EP14-A3 guideline does not provide any recommendations regarding the number of EQA material concentrations to be assessed (108). The results from our study, nevertheless, bring attention to the need for a consensus on the number of concentrations in which the EQA materials need to be tested, which is also discussed in the new IFCC recommendations (86, 110). The IFCC recommendations suggest, however, that each concentration should be evaluated separately as commutable or noncommutable (13). For reference materials, the primary function of which is to act as a calibrator, this is reasonable. For EQA organizations, however, it would be an advantage having pre-agreed numbers of EQA samples needed to be commutable for the EQA material overall to be considered commutable. This evolves from the fact that it is practically impossible for the EQA organization to assess every possible concentration of an EQA material. Additionally, even if the organization aimed to distribute similar concentrations of the EQA material in consecutive schemes, the concentrations would always vary somewhat. The results from our study, therefore, highlight the importance of performing studies on the guidelines' intended target group, in this case, the EQA providers' EQA material, as the results from commutability testing of reference material do not automatically apply for EQA materials.

The second issue is the noncommutability observed at higher concentrations. This finding might be caused by the abovementioned statistical limitation, or it may be

related to the concentration of the analyte. Although the study conducted by Liu et al. (127) did not specify whether the reference material was spiked, the EQA material produced by Noklus, and used in paper III, was spiked with human CRP and D(+)-glucose monohydrate, which was the only feature varying between the low and high concentrations of the EQA material. Supplementation, although with highly purified human analytes, is known to possibly cause impurities (150, 151).

In the study in paper III, the EQA material tested was a whole-blood EQA material customized by Noklus. The results from the study may, therefore, not be generalizable to other EQA materials produced in a different manner. The learnings from the study, however, can be useful for others when assessing the commutability of their material, either using the CLSI guideline step by step, like we did for CRP and Hb, or in a tailored procedure like that performed for glucose.

By distributing a commutable EQA material to the participants, harmonization among different analytes and laboratories can be assessed. In an international context, this is related to standardization and harmonization of laboratory results, whereas locally, it is related to the patients being treated with the potential benefit of making transitions between primary and secondary care safer and more trouble free. Finally, large lot variations have been observed for different analytes. When a commutable material is distributed to the participants, the variation seen between lots of the reagents of the MPs in the EQA material will also be present for different lots of clinical samples and can therefore have clinical implications.

5.3 Further research

Additional research on POC testing taking a broader perspective is required. When using POC rather than central laboratory testing, the advantages must outweigh the challenges. In a systematic review on GPs attitudes toward POC testing, some main concerns were identified, including the POC test's impact on clinical decisions and how it would affect clinical staff and workflow, issues related to cost, and regulation

and management of the quality of POC testing (152). Paper I and II in the current thesis addressed the quality management of POC testing, focusing on EQA. There is, however, little evidence on when to use IQC in POC testing, and further research could help determine when and how frequently IQC should be performed. Regarding quality management, the main focus has been on reducing errors in the analytical phase. Evaluation of the risk of errors in the entire testing process is, however, important. As seen in paper II of this thesis, which MP the participants used was a significant predictor for good quality. The POC MP will probably be even more important as they get more sophisticated, and functions for reducing both pre- and postanalytical errors are added. Research on whether choosing such MPs could affect the quality of POC testing would therefore be of great interest. Additionally, there is evidence that indicates serious failures in the way laboratory results are followed up, thus reflecting errors in the post- and post-postanalytical phase (153). How test results are evaluated and acted upon is key for best practice in diagnoses and management of patients and could be addressed in further research. However, as the nature of medical errors is multi-factorial, both people and system factors must, therefore, be addressed.

It is important that results obtained from observational studies are translated into meaningful improvement efforts in clinical practice. The research should ideally lead to the development of clinical guidelines, policy changes, or better patient outcomes. The outcome measures in current studies, including paper I and II in this thesis, can be a somewhat distant measure for actual patient outcome. Different clinical outcomes to be studied further, in which some naturally depend on the setting where POC testing is used, could be whether POC improves adherence to treatment, whether treatment is decided and started earlier, or if there is a reduction in reoperation and readmission rate and reduced incidence of complications or mortality rates. Economical outcomes of POC testing should also be considered because laboratory tests can already be performed in well-functional laboratories, and it is not given that rapid test results lead to savings. Potential economic benefits to be considered can, for instance, be reduced hospital admissions, better and less inappropriate drug use, reduction of staff and equipment, or improved life quality. There are currently few studies on POC testing

addressing safety, efficiency, sustainability, or the economics of primary care in this context (154). Well-designed studies aiming to detect whether quality in the POC testing process affects clinical and economical outcomes are, therefore, needed.

There has been improvement regarding the harmonization and standardization of laboratory test results. However, efforts to reduce interlaboratory variability should be prioritized. A key in achieving standardization is traceability and commutability. As seen in paper III of this thesis, commutability of the EQA material was determined for some analytes whereas for others, in our case, glucose, commutability was only determined for about half of the POC MPs. Further work on guidelines on how to assess commutability for different materials is, therefore, needed. In the case of EQA materials, further research on when an EQA material overall is commutable, i.e., how many concentrations of the material needs to be evaluated and how many of the evaluated samples must be commutable, needs to be assessed. In addition, how often an EQA material, produced in the same manner, should be reevaluated for commutability is of great importance because the EQA organization cannot assess commutability prior to every EQA scheme. There is also a need for further research on the commutability of both reference and EQA materials to achieve comparable results among MPs. Making the results comparable is important for developing clinical guidelines, which in turn are important for patient care related decisions.

5.4 Conclusion

In this thesis, we aimed to evaluate the effect of participating in an EQA organization and to evaluate the MPs and EQA material used to improve the participants' analytical quality.

The results from paper I and II, showed an increase in good participant performance and a decrease in poor participant performance depending on the number of years the participants took part in the Noklus quality improvement system. In addition, paper II showed that performing IQC weekly, performing the specific analysis ten or more

times weekly, using recommended MPs, and having laboratory-qualified operators predicted good participant performance in the EQAs. The results were, overall, similar for u-albumin, CRP, glucose, and Hb.

The results from paper III, showed that the whole-blood EQA material produced at Noklus was commutable for all CRP and Hb POC MPs and for about half of the glucose POC MPs used in primary health care in Norway. Consequently, a common reference value can be used to assess the participants' and Hb results. For glucose, all participant results were evaluated using method-specific target values. Because a commutable EQA material and reference values can be applied, the possibility to discover true analytical errors is improved, and consequently, possible errors in the treatment of patients and disease management can more likely be prevented. In addition, true systematic differences between the POC MPs can be discovered. This gives the EQA organizations' an opportunity to provide IVD manufactures feedback and also improve the guidance to participants on which MP to use. By having a commutable EQA material, the EQA organizations can monitor the degree of, and analytes in need of, standardization and harmonization.

In conclusion, systematic participation in quality management for POC testing in primary health care can improve the analytical quality of u-albumin, CRP, Hb, and glucose measurements. The whole-blood EQA material used was commutable for all CRP and Hb MPs and about half the glucose MPs used in primary healthcare in Norway. For these participants and MPs, both participant and MP evaluation can now be performed.

References

1. Comitato Elettrotecnico Italiano. VIM – International Vocabulary of Metrology [Internet]. Milano: Comitato Elettrotecnico Italiano; 2021. Available from: https://www.ceinorme.it/en/normazione-en/vim-en/vim-content-en.html
2. Bolann B.J. Riktig svar på biokjemiske analyser. Bergen: Fagbokforlaget; 2009.
3. Miller WG, Sandberg S. Quality Control of the Analytical Examination Process. In: Rifai N. Tietz Tetbook of Clinical Chemistry and Molecular Diagnostics. 6th ed: Elsevier; 2017.
4. Moore RE. Historical Perspective on the Clinical Diagnostic Laboratory. In: Molecular Diagnostics: For the Clinical Laboratorian. Springer Nature. [3-13]; 1997.
5. Plebani M. Does POCT reduce the risk of error in laboratory testing? Clin Chim Acta. 2009;404(1):59-64.
6. Belk WP, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. Am J Clin Pathol. 1947;17(11):853-61.
7. Wootton ID, King EJ. Normal values for blood constituents; inter-hospital differences. Lancet. 1953;1(6758):470-1.
8. Price CP, Smith I, Van den Bruel A. Improving the quality of point-of-care testing. Fam Pract. 2018;35(4):358-64.
9. Centers for Disease Control and Prevention. Influenza 2021. U.S. 2021. Available from: https://www.cdc.gov/flu/professionals/diagnosis/overview-testing-methods.html
10. Centers for Disease Control and Prevention. Test for Current Infection (Viral Test) [Internet]. Covid-19 2021. Atlanta: Centers for Disease Control and Prevention; 2021. Available from: https://www.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html
11. Junker R, Petersmann A, Luppa PB. The relevance of POCT in healthcare. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 9-17.
12. Luppa PB. Device classes. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 19-25.
13. O`Kane MJ. Point of Care Testing - Current and Emerging Quality Perspectives. Point care. 2014;13:1-5.
14. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on *in vitro* diagnostic medical devices, last amended by Commission Directive 2011/100/EU of 20 December 2011.
15. Spitzenberger F, Langer C, Gassner UM. Medical device legislation and POCT. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 251-60.
16. International Organization for Standardization. ISO 22870:2016: Point-of-care testing (POCT) - Requirements for quality and competence. Geneva: ISO; 2016.

17. Clinical & Laboratory Standards Institute. CLSI Guidelines [Internet]. Annapolis Junction: Clinical & Laboratory Standards Institute; 2021. Available from: https://clsi.org
18. Kost G. Principles and practice of point-of-care testing. Philadelphia: Lippincott Williams & Wilkins; 2002.
19. Jani IV, Peter TF. How point-of-care testing could drive innovation in global health. N Engl J Med. 2013;368(24):2319-24.
20. Price CP. Point of care testing. BMJ. 2001;322(7297):1285-8.
21. Irvin GL, 3rd, Molinari AS, Figueroa C, Carneiro DM. Improved success rate in reoperative parathyroidectomy with intraoperative PTH assay. Ann Surg. 1999;229(6):874-8; discussion 8-9.
22. Grieve R, Beech R, Vincent J, Mazurkiewicz J. Near patient testing in diabetes clinics: appraising the costs and outcomes. Health Technol Assess. 1999;3(15):1-74.
23. van Heyningen C, Watson ID, Morrice AE. Point-of-care testing outcomes in an emergency department. Clin Chem. 1999;45(3):437-8.
24. Dahler-Eriksen BS, Lauritzen T, Lassen JF, Lund ED, Brandslund I. Near-patient test for C-reactive protein in general practice: assessment of clinical, organizational, and economic outcomes. Clin Chem. 1999;45(4):478-85.
25. Fenwick EA, Briggs AH, Hawke CI. Management of urinary tract infection in general practice: a cost-effectiveness analysis. Br J Gen Pract. 2000;50(457):635-9.
26. Sawicki PT. A structured teaching and self-management program for patients receiving oral anticoagulation: a randomized controlled trial. Working Group for the Study of Patient Self-Management of Oral Anticoagulation. JAMA. 1999;281(2):145-50.
27. Fraunberger P, Gruber S, Amiet F, Fiedler M, Vaubourdolle M, Beneteau-Burnat B, et al. Quality assurance in POCT – A cross-country comparison. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 345-74.
28. Haldrup S, Thomsen RW, Bro F, Skov R, Bjerrum L, Sogaard M. Microbiological point of care testing before antibiotic prescribing in primary care: considerable variations between practices. BMC Fam Pract. 2017;18(1):9.
29. Lundberg GD. Acting on significant laboratory results. JAMA. 1981;245(17):1762-3.
30. Hofer TP, Kerr EA, Hayward RA. What is an error? Eff Clin Pract. 2000;3(6):261-9.
31. Sunshine JE, Meo N, Kassebaum NJ, Collison ML, Mokdad AH, Naghavi M. Association of Adverse Effects of Medical Treatment With Mortality in the United States: A Secondary Analysis of the Global Burden of Diseases, Injuries, and Risk Factors Study. JAMA Netw Open. 2019;2(1):e187041.
32. Plebani M. The detection and prevention of errors in laboratory medicine. Ann Clin Biochem. 2010;47(Pt 2):101-10.
33. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem. 2007;53(7):1338-42.

34. Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. Clin Chem. 1997;43(8 Pt 1):1348-51.
35. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324(6):377-84.
36. Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem. 2002;48(5):691-8.
37. Nichols JH. How to achieve quality for POCT through risk management. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg. p. 393-409.
38. U.S. Food & Drug Administration. Useul tips to Increase Accuracy and Reduce Errors in Test Results from Glucose Meters [Internet]. Washington D.C: U.S. Food and Drug Administration; 2015. Available from: http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/TipsandArticlesonDeviceSafety/ucm109519.htm39.
40. Meier FA, Jones BA. Point-of-care testing error: sources and amplifiers, taxonomy, prevention strategies, and detection monitors. Arch Pathol Lab Med. 2005;129(10):1262-7.
41. Fraser C. Biologic Variation: Principles and Practice: Westgard QC; Available from: https://www.westgard.com/guest19.htm
42. Norsk selskap for medisinsk biokjemi. Nasjonal brukerhåndbok i Medisinsk Biokjemi [Internet]. Norsk selskap for medisinsk biokjemi; 2021. Available from: https://brukerhandboken.no
43. Aarsand AK F-CP, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, Jonker N, Minchinela J, Simon M, Braga F, Perich C, Boned B, Roraas T, Marques-Garcia F, Carobene A, Aslan B, Barlett WA, Sandberg S. The EFLM Biological Variation Database (2021). Available from: https://biologicalvariation.eu
44. Clinical and Laboratory Standards Institute. EP21-A: Estimation Of Total Analytical Error For Clinical Laboratory Methods; Approved Guideline. Pennsylvania: CLSI; 2003.
45. Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on quality specifications in laboratory medicine, 25-26 April 1999. Scand J Clin Lab Invest. 1999;59(7):475-6.
46. Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. Clin Chem Lab Med. 2015;53(6):829-32.
47. Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T, et al. Recommendations for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. Clin Chem. 2006;52(1):5-18.
48. National Institues of Health. Recommendations for improving cholesterol measurement: a report from the Laboratory Standardization Panel of the National Cholesterol Education Program. Bethesda: National Institutes of Health; 1990.

49. Horvath AR, Bossuyt PM, Sandberg S, John AS, Monaghan PJ, Verhagen-Kamerbeek WD, et al. Setting analytical performance specifications based on outcome studies - is it possible? Clin Chem Lab Med. 2015;53(6):841-8.
50. Reason J. Human error: models and management. BMJ. 2000;320(7237):768-770.
51. Clinical and Laboratory Standards Institute. EP23-A: Laboratory Quality Control Based on Risk Management; Approved Guideline. Pennsylvania: CLSI; 2011.
52. Nichols JH. Quality in point-of-care testing. Expert Rev Mol Diagn. 2003;3(5):563-72.
53. Holt H, Freedman DB. Internal quality control in point-of-care testing: where's the evidence? Ann Clin Biochem. 2016;53(Pt 2):233-9.
54. Burnett D, Ceriotti F, Cooper G, Parvin C, Plebani M, Westgard J. Collective opinion paper on findings of the 2009 convocation of experts on quality control. Clin Chem Lab Med. 2010;48(1):41-52.
55. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem. 2011;57(12):1670-80.
56. Clinical and Laboratory Standards Institute. GP27-A2: Using Proficiency Testing to Improve the Clincal Laboratory; Approved Guideline. Pennsylvania: CLSI; 2007.
57. International Organization for Standardization. ISO/IEC 17043:2010: Conformity assessment: general requirements for proficiency testing. Geneva: ISO;2010.
58. Jones GRD, Albarede S, Kesseler D, MacKenzie F, Mammen J, Pedersen M, et al. Analytical performance specifications for external quality assessment - definitions and descriptions. Clin Chem Lab Med. 2017;55(7):949-55.
59. Blirup-Jensen S, Johnson AM, Larsen M, Proteins ICoP. Protein standardization V: value transfer. A practical protocol for the assignment of serum protein values from a Reference Material to a Target Material. Clin Chem Lab Med. 2008;46(10):1470-9.
60. Broughton PM, Eldjarn L. Methods of assigning accurate values to reference serum. Part 1. The use of reference laboratories and consensus values, with an evaluation of a procedure for transferring values from one reference serum to another. Ann Clin Biochem. 1985;22 (Pt 6):625-34.
61. The Joint Committee for Traceability in Laboratory Medicine (JCTLM) [Internet]. Database of higher-order refernce materials, measurement methods/procedures and services. Sevres Cedex: Burea International des Poids et Mesures; 2021. Avaliable from: http://www.bipm.org/jctlm/
62. Xavier A. How is poor performance defined among EQA organisations? [Internet]. Toulouse: EQALM;2014; Available from: http://www.eqalm.org/site//2014/1-2_X_Albe_Poor_Performance.pdf
63. Centers for Disease Control and Prevention. Current CLIA regulations 2021. [Internet]. Atlanta: Centers for Disease Control and Prevention; 2021. Available from: http://wwwn.cdc.gov/clia/regs/toc.aspx

64. Fraser CG, Kallner A, Kenny D, Petersen PH. Introduction: strategies to set global quality specifications in laboratory medicine. Scand J Clin Lab Invest. 1999;59(7):477-8.
65. Ross JW, Fraser MD. Analytical goals developed from the inherent error of medical tests. Clin Chem. 1993;39(7):1481-93; discussion 93-4.
66. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med. 2015;53(6):833-5.
67. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta. 2003;327(1-2):25-37.
68. Miller WG, Myers GL, Ashwood ER, Killeen AA, Wang E, Thienpont LM, et al. Creatinine measurement: state of the art in accuracy and interlaboratory harmonization. Arch Pathol Lab Med. 2005;129(3):297-304.
69. Miller WG, Myers GL, Ashwood ER, Killeen AA, Wang E, Ehlers GW, et al. State of the art in trueness and interlaboratory harmonization for 10 analytes in general clinical chemistry. Arch Pathol Lab Med. 2008;132(5):838-46.
70. Little RR, Rohlfing CL, Sacks DB, National Glycohemoglobin Standardization Program Steering C. Status of hemoglobin A1c measurement and goals for improvement: from chaos to order for improving diabetes care. Clin Chem. 2011;57(2):205-14.
71. Miller WG. Estimating glomerular filtration rate. Clin Chem Lab Med. 2009;47(9):1017-9.
72. Kroll MH, Chesler R. Effect of serum lyophilization on the rate constants of enzymatic methods for measuring cholesterol. Clin Chem. 1990;36(3):534-7.
73. Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem. 2006;52(4):553-4.
74. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev. 2007;28(4):139-47.
75. Stavelin A, Sandberg S. Harmonization activities of Noklus - a quality improvement organization for point-of-care laboratory examinations. Clin Chem Lab Med. 2018;57(1):106-14.
76. Noklus. The Norwegian quality improvement of primary health care laboratories [Internet]. Bergen: Noklus; 2021. Available from: https://www.noklus.no
77. International Organization for Standardization. ISO 9001:2015: Quality management systems - Requirements. Geneva: ISO; 2015
78. International Organization for Standardization. ISO 13528:2015: Statistical methods for use in proficiency testing by interlaboratory comparison. Geneva: ISO; 2015.
79. International Organization for Standardization. ISO 17511:2020: In vitro diagnostic medical devices - Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. Geneva: ISO; 2020.

80. Vesper HW, Thienpont LM. Traceability in laboratory medicine. Clin Chem. 2009;55(6):1067-75.
81. Panteghini M. Traceability, reference systems and result comparability. Clin Biochem Rev. 2007;28(3):97-104.
82. Sturgeon CM, Berger P, Bidart JM, Birken S, Burns C, Norman RJ, et al. Differences in recognition of the 1st WHO international reference reagents for hCG-related isoforms by diagnostic immunoassays for human chorionic gonadotropin. Clin Chem. 2009;55(8):1484-91.
83. Gantzer ML, Miller WG. Harmonisation of Measurement Procedures: how do we get it done? Clin Biochem Rev. 2012;33(3):95-100.
84. Miller WG, Tate JR, Barth JH, Jones GR. Harmonization: the sample, the measurement, and the report. Ann Lab Med. 2014;34(3):187-97.
85. International Consortium for Harmonization of Clinical Laboratory Results. About [Internet]. ICHCLR. Available from: https://www.harmonization.net/about/
86. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. Clin Chem. 2018;64(3):447-54.
87. Bunk DM. Reference materials and reference measurement procedures: an overview from a national metrology institute. Clin Biochem Rev. 2007;28(4):131-7.
88. Whicher JT, Ritchie RF, Johnson AM, Baudner S, Bienvenu J, Blirup-Jensen S, et al. New international reference preparation for proteins in human serum (RPPHS). Clin Chem. 1994;40(6):934-8.
89. Merlini G, Blirup-Jensen S, Johnson AM, Sheldon J, Zegers I, Proteins ICoP. Standardizing plasma protein measurements worldwide: a challenging enterprise. Clin Chem Lab Med. 2010;48(11):1567-75.
90. Baudner S, Bienvenu J, Bilrup-Jensen S, Carlstrom A, Johnson AM, Milford Ward A, et al. The certification of a matrix reference material for immunochemical measurement of 14 human serum proteins: CRM470, EUR 15243 EN. Luxembourg: Commission of the European Communities; 1993.
91. Shrivastava AK, Singh HV, Raizada A, Singh SK. C-reactive protein, inflammation and coronary heart disease. The Egyptian Heart Journal. 2015;67(2):89-97.
92. Thompson D, Pepys MB, Wood SP. The physiological structure of human C-reactive protein and its comple with phosphocholine. Structure. 1999;7(2):169-77.
93. Rzychon M, Zegers I, Schimmel H. Analysis of the physicochemical state of C-reactive protein in different preparations including 2 certified reference materials. Clin Chem. 2010;56(9):1475-82.
94. Joint Research Centre Institute for Reference Materials and Measurements. The certification of the mass concentration of C-reactive protein in human serum - Certified reference material ERM®-DA474/IFCC 2009. Geel: European Commission JRC-IRMM. Available from: https://op.europa.eu/en/publication-detail/-/publication/7ed403a7-446f-4eb6-ac2d-badb16a62220/language-en

95. National Institute for Biological Standards and Control. WHO International Standard Human C-reactive protein [Internet]. South Mimms: NIBSC; 2008. Available from: https://www.nibsc.org/products/brm_product_catalogue/detail_page.aspx?catid=85/506
96. Hones J, Muller P, Surridge N. The Techhnology Behind Glucose Meters: Test strips. Diabetes Technol Ther. 2008;10:10-26.
97. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, Kulpmann WR, et al. Approved IFCC recommendation on reporting results for blood glucose: International Federation of Clinical Chemistry and Laboratory Medicine Scientific Division, Working Group on Selective Electrodes and Point-of-Care Testing (IFCC-SD-WG-SEPOCT). Clin Chem Lab Med. 2006;44(12):1486-90.
98. Stockl D, Reinauer H. Candidate reference methods for determining target values for cholesterol, creatinine, uric acid, and glucose in external quality assessment and internal accuracy control. I. Method setup. Clin Chem. 1993;39(6):993-1000.
99. Thienpont LM, Leenheer AP, Stockl D, Reinauer H. Candidate reference methods for determining target values for cholesterol, creatinine, uric acid, and glucose in external quality assessment and internal accuracy control. II. Method transfer. Clin Chem. 1993;39(6):1001-6.
100. National institute of Standards and Technology (2006). Standard Reference Material 917b - d-glucose (Dextrose). Avaliable from: https://www-s.nist.gov/srmors/certificates/archive/917b.pdf
101. Zijlstra WG. Standardisation of haemoglobinometry: History and new challenges. Comparative Haematology International. 1997;7:125-32.
102. Britannica [Internet]. London: Encyclopædia Britannica. Hemoglobin [cited 2021 Apr 1]. Available from: https://www.britannica.com/science/hemoglobin
103. Tentori L, Salvati AM. Hemoglobinometry in human blood. Methods Enzymol. 1981;76:707-15.
104. Publications Office of the EU. Certified Reference Material - BCR - 522. Bovine Blood Lysate. Brussels; 1994.
105. Christenson RH, Duh SH, Apple FS, Bodor GS, Bunk DM, Panteghini M, et al. Toward standardization of cardiac troponin I measurements part II: assessing commutability of candidate reference materials and harmonization of cardiac troponin I assays. Clin Chem. 2006;52(9):1685-92.
106. International Organization for Standardization. ISO Guide 35:2017: Reference materials - Guidance for characterization and assessment of homogeneity and stability. Geneva: ISO; 2017.
107. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS, et al. Executive summary: guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem. 2011;57(6):793-8.
108. Clinical and Laboratory Standards Institute. EP14-A3: Evaluation of Commutability of Processed Samples; Approved Guideline. Pennsylvania: CLSI; 2014.

109. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC Working Group Recommendations for Assessing Commutability Part 3: Using the Calibration Effectiveness of a Reference Material. Clin Chem. 2018;64(3):465-74.
110. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. Clin Chem. 2018;64(3):455-64.
111. Sohn AJ, Hickner JM, Alem F. Use of Point-of-Care Tests (POCTs) by US Primary Care Physicians. J Am Board Fam Med. 2016;29(3):371-6.
112. Hardy V, Thompson M, Alto W, Keppel GA, Hornecker J, Linares A, et al. Exploring the barriers and facilitators to use of point of care tests in family medicine clinics in the United States. BMC Fam Pract. 2016;17(1):149.
113. Shaw J. The dark side of point-of-care testing. Clin Biochem. 2017;50(9):466-7.
114. Herd G, Musaad SM. Clinical governance and point-of-care testing at health provider level. N Z Med J. 2015;128(1417):41-6.
115. Huddy JR, Ni MZ, Barlow J, Majeed A, Hanna GB. Point-of-care C reactive protein for the diagnosis of lower respiratory tract infection in NHS primary care: a qualitative study of barriers and facilitators to adoption. BMJ Open. 2016;6(3):e009959.
116. Stavelin A, Sandberg S. Essential aspects of external quality assurance for point-of-care testing. Biochem Med (Zagreb). 2017;27(1):81-5.
117. Clinical and Laboratory Standards Institute. EP14-A2: Evaluation of Matrix Effects; Approved Guideline. Pennsylvania: CLSI; 2005
118. Weykamp C. HbA1c: a review of analytical and clinical aspects. Ann Lab Med. 2013;33(6):393-400.
119. Husøy AM. Blodprøvetaking i praksis. Oslo: Cappelen Damm Akademiske; 2012.
120. Beetham R, White P, Riches P, Bullock D, MacKenzie F, Group UNPASA. Use of CRM 470/RPPHS has not achieved true consensus for ceruloplasmin measurement. Clin Chem. 2002;48(12):2293-4.
121. Nationale Institute of Standards and Technology. Standard Reference Material 965b - glucose in frozen human serum [Internet]. Gaithersburg: National Institute of Standards & Technology; 2016. Avaliable from: https://www-s.nist.gov/srmors/certificates/965B.pdf
122. Kristensen GB, Christensen NG, Thue G, Sandberg S. Between-lot variation in external quality assessment of glucose: clinical importance and effect on participant performance evaluation. Clin Chem. 2005;51(9):1632-6.
123. Stavelin A, Riksheim BO, Christensen NG, Sandberg S. The Importance of Reagent Lot Registration in External Quality Assurance/Proficiency Testing Schemes. Clin Chem. 2016;62(5):708-15.
124. Stockl D, Stepman HC, Van Houcke SK, Thienpont LM. Importance of sample-related effects for commutability testing according to the EP14 protocol. Clin Chim Acta. 2010;411(17-18):1378-9.

125. Roraas TH, Van Houcke SK, Stockl D, Thienpont LM. Statistical power for commutability testing in the presence of random, sample-related effects by use of the EP14 protocol. Clin Chim Acta. 2012;413(19-20):1710-1.
126. Zhang S, Zeng J, Zhang C, Li Y, Zhao H, Cheng F, et al. Commutability of possible external quality assessment materials for cardiac troponin measurement. PLoS One. 2014;9(7):e102046.
127. Liu H, Wong L, Yong S, Liu Q, Teo TL, Lee TK, et al. Commutable whole blood reference materials for hemoglobin A1c validated on multiple clinical analyzers. Clin Chem Lab Med. 2019;57(5):648-58.
128. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias Between a Reference Material and Clinical Samples. Clin Chem. 2018.
129. Yan Y, Han B, Zhao H, Ma R, Wang J, Wang D, et al. Commutability of external quality assessment materials for serum sodium and potassium measurements. Clin Chem Lab Med. 2019;57(4):465-75.
130. Delatour V, Liu Q, Vesper HW, Commutability L-LWGo. Commutability Assessment of External Quality Assessment Materials with the Difference in Bias Approach: Are Acceptance Criteria Based on Medical Requirements too Strict? Clin Chem. 2016;62(12):1670-1.
131. Korzun WJ, Nilsson G, Bachmann LM, Myers GL, Sakurabayashi I, Nakajima K, et al. Difference in bias approach for commutability assessment: application to frozen pools of human serum measured by 8 direct methods for HDL and LDL cholesterol. Clin Chem. 2015;61(8):1107-13.
132. Lacobbucci G. Covid-19: Government is criticised for «scandalous» £10bn spent on test and trace programme. BMJ. 2020.
133. Libeer JC, Baadenhuijsen H, Fraser CG, Petersen PH, Ricos C, Stockl D, et al. Characterization and classification of external quality assessment schemes (EQA) according to objectives such as evaluation of method and participant bias and standard deviation. External Quality Assessment (EQA) Working Group A on Analytical Goals in Laboratory Medicine. Eur J Clin Chem Clin Biochem. 1996;34(8):665-78.
134. Sciacovelli L, Zardo L, Secchiero S, Plebani M. Quality specifications in EQA schemes: from theory to practice. Clin Chim Acta. 2004;346(1):87-97.
135. Shephard MD, Gill JP. Results of an innovative education, training and quality assurance program for point-of-care HbA1c testing using the Bayer DCA 2000 in Australian Aboriginal Community Controlled Health Services. Clin Biochem Rev. 2003;24(4):123-30.
136. Morandi PA, Deom A, Kesseler D, Cohen R. Retrospective analysis of 55,769 HbA1c EQA results obtained from professional laboratories and medical offices participating in surveys organized by two European EQA centers over a nine-year period. J Clin Lab Anal. 2011;25(5):337-43.
137. Solvik UO, Roraas T, Christensen NG, Sandberg S. Diagnosing diabetes mellitus: performance of hemoglobin A1c point-of-care instruments in general practice offices. Clin Chem. 2013;59(12):1790-801.

138. Mayo E. The human problems of an industrial civilization. New York: Routledge; 2004.
139. Bietenbeck A, Geilenkeuser WJ, Klawonn F, Spannagl M, Nauck M, Petersmann A, et al. External quality assessment schemes for glucose measurements in Germany: factors for successful participation, analytical performance and medical impact. Clin Chem Lab Med. 2018;56(8):1238-50.
140. The King's Fund. Understanding Pressure in General Practice [Internet]. London: The King's Fund; 2016. Available from: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Understanding-GP-pressures-Kings-Fund-May-2016.pdf
141. NHS England. Five Year Forward View [Internet]. Redditch: NHS England; 2014. Available from: https://www.england.nhhs.uk/wp-contrnt/uploads/2014/10/5yfv-web.pdf
142. Tollanes MC, Jenum AK, Berg TJ, Lovaas KF, Cooper JG, Sandberg S. Availability and analytical quality of hemoglobin A1c point-of-care testing in general practitioners' offices are associated with better glycemic control in type 2 diabetes. Clin Chem Lab Med. 2020;58(8):1349-56.
143. Scandinavian evaluation of laboratory equipment for primary health care. SKUP. Bergen: SKUP; 2021. Available from: http://www.skup.nu/
144. Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine chemistry analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. Arch Pathol Lab Med. 1998;122(7):587-608.
145. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem. 1998;31(6):449-57.
146. Meng Q, Zhou W, Zhang C, Zeng J, Zhao H, Zhang T, et al. Serum triglyceride measurements: the commutability of reference materials and the accuracy of results. Clin Chem Lab Med. 2017;55(9):1284-90.
147. Zegers I, Beetham R, Keller T, Sheldon J, Bullock D, MacKenzie F, et al. The importance of commutability of reference materials used as calibrators: the example of ceruloplasmin. Clin Chem. 2013;59(9):1322-9.
148. Yan Y, Pu Y, Long Q, Zhang J, Zhang T, Zhou W, et al. Commutability of external quality assessment materials for serum magnesium and calcium measurements. Scand J Clin Lab Invest. 2019;79(6):404-11.
149. van der Hagen EAE, Weykamp C, Sandberg S, Stavelin AV, MacKenzie F, Miller WG. Feasibility for aggregation of commutable external quality assessment results to evaluate metrological traceability and agreement among results. Clin Chem Lab Med. 2020;59(1):117-25.
150. Howanitz JH. Review of the influence of polypeptide hormone forms on immunoassay results. Arch Pathol Lab Med. 1993;117(4):369-72.
151. Rej R. Accurate enzyme activity measurements. Two decades of development in the commutability of enzyme quality control materials. Arch Pathol Lab Med. 1993;117(4):352-64.

152. Jones CH, Howick J, Roberts NW, Price CP, Heneghan C, Pluddemann A, et al. Primary care clinicians' attitudes towards point-of-care blood testing: a systematic review of qualitative studies. BMC Fam Pract. 2013;14:117.
153. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. J Gen Intern Med. 2012;27(10):1334-48.
154. Phillips CB, Pearce CM, Hall S, Travaglia J, de Lusignan S, Love T, et al. Can clinical governance deliver quality improvement in Australian general practice and primary care? A systematic review of the evidence. Med J Aust. 2010;193(10):602-7.

I

Tone Bukve*, Thomas Røraas, Berit Oddny Riksheim, Nina Gade Christensen and Sverre Sandberg

Point-of-care urine albumin in general practice offices: effect of participation in an external quality assurance scheme

Abstract

Background: The Norwegian Quality Improvement of Primary Care Laboratories (Noklus) offers external quality assurance (EQA) schemes (EQASs) for urine albumin (UA) annually. This study analyzed the EQA results to determine how the analytical quality of UA analysis in general practice (GP) offices developed between 1998 (n=473) and 2012 (n=1160).

Methods: Two EQA urine samples were distributed yearly to the participants by mail. The participants measured the UA of each sample and returned the results together with information about their instrument, the profession and number of employees at the office, frequency of internal quality control (IQC), and number of analyses per month. In the feedback report, they received an assessment of their analytical performance.

Results: The number of years that the GP office had participated in Noklus was inversely related to the percentage of "poor" results for quantitative but not semiquantitative instruments. The analytical quality improved for participants using quantitative instruments who received an initial assessment of "poor" and who subsequently changed their instrument. Participants using reagents that had expired or were within 3 months of the expiration date performed worse than those using reagents that were expiring in more than 3 months.

Conclusions: Continuous participation in the Noklus program improved the performance of quantitative UA analyses at GP offices. This is probably in part attributable to the complete Noklus quality system, whereby in addition to participating in EQAS, participants are visited by laboratory consultants who examine their procedures and provide practical advice and education regarding the use of different instruments.

Keywords: external quality assurance schemes; point-of-care testing; primary health care; urine albumin.

DOI 10.1515/cclm-2014-0483
Received May 5, 2014; accepted July 8, 2014; previously published online August 7, 2014

***Corresponding author: Tone Bukve,** Noklus, P.O. Box 6165, 5892 Bergen, Norway, Phone: +47-95818805, Fax: +47-55979510, E-mail: buto@hotmail.no; and The Norwegian Quality Improvement of Primary Care Laboratories (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway

Thomas Røraas, Berit Oddny Riksheim and Nina Gade Christensen: The Norwegian Quality Improvement of Primary Care Laboratories (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway

Sverre Sandberg: The Norwegian Quality Improvement of Primary Care Laboratories (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Faculty of Medicine and Dentistry, Department of Global Health and Primary Care, University of Bergen, Bergen, Norway

Introduction

Diabetic nephropathy occurs in 20%–40% of patients with diabetes mellitus [1]. Urine albumin (UA) is a well-established marker for chronic kidney disease with or without diabetes and plays an important role in its early diagnosis, assessment, and treatment [1–3]. UA has also been shown to be an important marker for cardiovascular disease [4–7]. Measurements of albumin excretion may be performed in timed collections, early morning urine, or random urine collections. It is, however, recommended that early morning void and expression of results as albumin/creatinine ratio (ACR) are used [8]. For clinicians to be able to detect UA and begin to intervene when necessary, they need an analytically reliable result. Quality assurance of UA is therefore essential to ensure that laboratories provide reliable results. The Norwegian Quality Improvement of Primary Care Laboratories (Noklus) is responsible for external quality assurance (EQA) in Norway and has since 1992 offered EQA schemes (EQASs) for most of the analytes analyzed in primary health care. UA is part of the EQA program at Noklus, as well as in several other EQA programs worldwide [9]. It has been reported that the

EQAS is important for improving analytical quality [10]; however, there is little documented information or evidence supporting this assumption.

The present study analyzed the results of the EQAS, with the aim of determining how the analytical quality of UA analysis in the general practice (GP) offices in Norway developed between 1998 and 2012 in relation to the length of the participants' involvement in EQA and to evaluate what factors in the GP offices were associated with good analytical quality.

Materials and methods

Participants

During the period 1998–2012, 15 UA EQASs were distributed to GP offices in Norway; the total percentage of GP offices participating voluntarily in the Noklus quality assurance system, including participation in EQASs for the constituents that they analyze, is 99.7%. The GPs buy their own instruments; the point-of-care (POC) semiquantitative instruments used between 1998 and 2012 were Clinitek Status+, Clinitek Status, Clinitek 50, and Clinitek 100 (Siemens Healthcare Diagnostics), and MICRAL-TEST (Roche Diagnostics), whereas the quantitative instruments were Afinion AS100 Analyzer and NycoCard U-Albumin (Axis-Shield PoC), DCA 2000 Analyzer, DCA 2000+ Analyzer, and DCA Vantage Analyzer (Siemens Healthcare Diagnostics), HemoCue Urine Albumin and HemoCue Albumin 201 (HemoCue), and QuikRead U-ALB and QuikRead101 (Orion Diagnostica).

EQA survey

The EQA material was obtained by mixing samples from volunteers with normal or increased excretion of UA to achieve appropriate concentrations of UA. Noklus collected the samples before each survey and stored them at −80 °C. A few of the samples were stored for up to 2 years before they were circulated. The samples were sterile filtered prior to their circulation. They were then distributed at room temperature by mail, and the participants were instructed to keep them refrigerated after receiving them. Each GP office received two EQA samples containing albumin at two different concentrations. During the study period (i.e., 1998–2012), the concentration of albumin ranged from about 15 to 130 mg/L, with the mean for the low- and high-concentration samples being 27 and 78 mg/L, respectively. The participants analyzed each EQA sample on two different days (1×2 design) when quantitative methods were used and on one day (1×1 design) when semiquantitative methods were used.

The participants returned the analytical results within 6 days of the mailing day. The EQA samples were, according to the ISO 13528, stable and homogeneous during this period [9, 11]. The participants also provided information about the POC instrument, the lot number and expiration date of the reagents used, the numbers of general practitioners and their coworkers at the GP office, the education level of the person who usually performed the UA analysis, how often internal quality control (IQC) was performed, and the number of UA analyses performed monthly. Only results concerning UA are reported in the present article, as only DCA instruments were used to measure ACR throughout the whole study period.

Assessment

Trueness was defined in this study as closeness of the mean of duplicate measurements and the method-specific target value; precision was calculated as the difference between duplicate measurements; and accuracy was defined as closeness of the agreement between a measured quantity value and the method-specific target value. These definitions are modified from those in the International Vocabulary of Metrology [12].

Although a candidate reference method has been suggested [13], the combination of no reference method being accepted in the period of the study and the freezing of the EQA samples, which might lead to matrix effects and non-commutability between the methods [14], method-specific target values were used. For the quantitative methods, the target value was a "trimmed" median, calculated for instrument groups containing ≥20 participants after excluding outliers that were more than ±3 SDs from the median. The semiquantitative instruments reported their results in one of four possible categories, and the categories with >30% of the semiquantitative answers were set as the target value.

For quantitative instruments, each EQA sample was analyzed twice, and the participants received assessments ("very good", "acceptable", and "poor") concerning trueness and precision for each of the two levels. For semiquantitative instruments, the samples were analyzed once, and feedback on accuracy ("good", "questionable", and "poor") was provided for each of the two levels. Table 1 lists the quality specifications used by Noklus for quantitative and semiquantitative methods. The quality specifications are much stricter for quantitative methods than for semiquantitative methods. Examples on how these quality specifications are calculated are given in online Supplemental Data, Table 1, that accompanies the article at http://www.degruyter.com/view/j/cclm.2015.53.issue-1/issue-files/cclm.2015.53.issue-1.xml.

When analyzing results from both EQA samples together, the following categories were used for quantitative methods:

1. "Very good" trueness/precision: Participants obtaining "very good" trueness/precision assessments on both EQA samples.
2. "Acceptable" trueness/precision: Participants that obtained "acceptable" trueness/precision assessment on one EQA sample and "acceptable" or "very good" trueness/precision on the other.
3. "Poor" trueness/precision: All the other participants.

Similarly, the following categories were used when evaluating participants using semiquantitative instrument:

1. "Good" accuracy: Participants obtaining "good" accuracy assessments on both EQA samples.
2. "Questionable" accuracy: Participants that obtained "questionable" accuracy on one EQA sample and "questionable" or "very good" trueness on the other.
3. "Poor" accuracy: All the other participants.

The GP offices are defined as "new", both when they participate in the EQASs for the first time and when they change from a semiquantitative to a quantitative instrument; otherwise, they are defined as "old".

Table 1 Noklus quality specifications for assessing trueness, precision, and accuracy in the UA EQA.[a]

	Quality specifications
Quantitative instruments	
Target interval	Within ±2 mg/L of the method-specific target value
Trueness	
Very good	Mean value of duplicate measurements within ±7%
Acceptable	Mean value of duplicate measurements within target ±15%
Poor	Mean value of duplicate measurements outside ±15%
Precision	
Very good	Difference between duplicate measurements: ≤target value lower limit for "very good" trueness
Acceptable	Difference between duplicate measurements: values between "very good" and "poor" precision
Poor	Difference between duplicate measurements: target value lower limit for "acceptable" trueness
Semiquantitative instruments	
Accuracy	
Good	The categories with >30% of the answers
Questionable	The "neighboring" categories
Poor	The remaining categories

[a]Examples on how this is calculated is given in Supplemental data Table 1.

The influence of the expiration date on the participants' assessments was evaluated. Participants using the Afinion instruments were excluded from the group using expired reagents because this instrument does not accept reagents with an expired date.

Statistics

The two-sided χ^2-test and Fisher's exact test were used to determine whether there were significant differences between the odds ratios (ORs). The level of statistical significance was set at $p<0.05$. Testing for binominal proportions was used to explore the effects of changing an instrument and to determine any significant changes in practice characteristics between 1998 and 2012.

Logistic regression was applied to each EQA survey. Dependent variables were the trueness and precision of the assessments, and the independent variables were the instrument used, profession of the person who performed the analysis, frequency of IQC, number of GP employees at the office (until 2006), and number of analyses per month (from 2009). The data were analyzed using SPSS version 19.0.

Results

The percentage of GP offices analyzing UA and participating in the UA Noklus EQAS, relative to the total number of Norwegian GP offices, increased from 26% (473/1809) in 1998 to 68% (1163/1723) in 2012. Semiquantitative instruments were used by 76% in 1998 but only 9% in 2012 (see Supplemental Data, Figure 1). Figures 1A and 2A show the percentage of GP offices that achieved "very good"/"good" or "poor" assessments relative to the year of the EQA survey. From 1998 to 2003, the percentage of GP offices with assessments "very good" trueness (quantitative) and "good" accuracy (semiquantitative) decreased from 92% to 67% and from 86% to 58% for quantitative and semiquantitative instruments, respectively. After 2003, the percentage with assessments of "very good" trueness or "good" accuracy increased, then stabilized at 75%–90%.

Figures 1B and 2B show the percentage of GP offices obtaining "very good"/"good" or "poor" assessments relative to the number of years they have participated in the Noklus EQA program. The percentage of GP offices obtaining an assessment of "poor" trueness using a quantitative instrument decreased the longer they had participated in the EQAS. A similar decrease was not seen for semiquantitative instruments.

Comparing "old" to "new" participants using quantitative instruments, the OR for being assessed as "very good" trueness rather than "poor" in general increased with the number of years that the GP office had participated in the EQAS [i.e., comparison of participating for the third and ninth times yielded ORs of 1.5, 95% confidence interval (CI) 1.1–2.1, and 4.2, 95% CI: 2.2–7.8, respectively]. Regarding the precision, significantly more "old" participants compared to "new" participants were assessed as "very good" compared with "poor" (OR, 1.5, 95% CI, 1.2–1.8). "Old" participants using only a quantitative DCA instrument (DCA 2000, DCA 2000+, or DCA Vantage) were analyzed separately to eliminate possible effects of changing instruments. "Old" participants performed significantly better than "new" participants concerning trueness (OR, 2.0, 95% CI, 1.3–3.3)] but not precision.

The mean values of the instruments used in surveys have differed, as exemplified by the survey in 2012 where the UA at the low concentration was 33.6 mg/L (95%

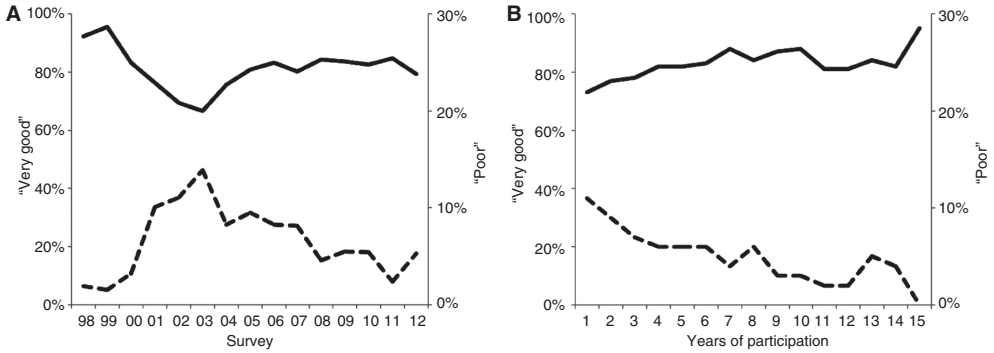


Figure 1 Quantitative instruments: percentage of participants that achieved "very good" assessments (solid black line) and "poor" assessments (dashed black line) relative to the year of the EQA survey (A) or the number of years of participation in Noklus (B).
Year 1 includes GP offices participating for the first time using quantitative instruments (n=1249). Year 15 includes GP offices participating in 15 EQASs using quantitative instruments (n=19).

CI, 33.4–33.8) for the DCA instruments, compared with 30.8 mg/L (95% CI, 30.6–31.1) and 34.7 mg/L (95% CI, 34.4–34.9) for the Afinion and HemoCue instruments (HemoCue Urine Albumin and HemoCue Albumin 201), respectively.

Changing instruments

Participants using quantitative instruments that received an assessment of "poor" trueness and subsequently changed their instrument received significantly fewer "poor" assessments the following year, regardless of whether they changed to another type of the same instrument or to a different type of quantitative instrument, compared with those who did not change instrument. When participants with an initial assessment of "very good" trueness changed to a different instrument, significantly fewer participants received "very good" trueness assessments, compared with when the participants did not change instrument (Table 2). Similar results were not found for participants using semiquantitative instruments (results not shown).

Several practice characteristics changed significantly between 1998 and 2012. For example, the

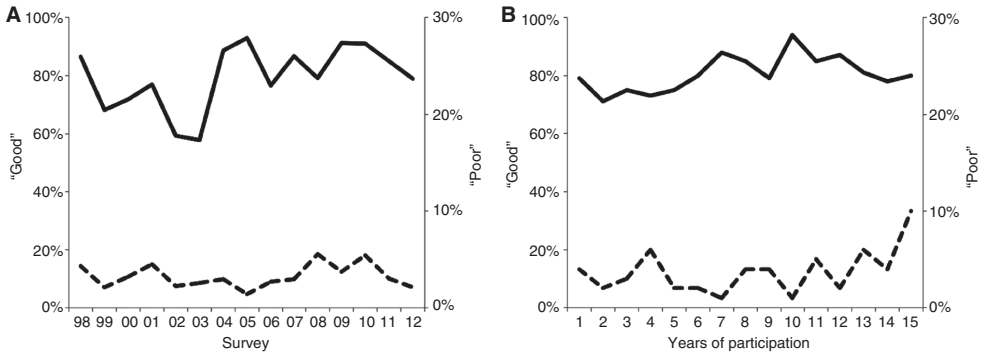


Figure 2 Semiquantitative instruments: percentage of participants that achieved "good" assessments (solid black line) and "poor" assessments (dashed black line) relative to the year of the EQA survey (A) or the number of years of participation in Noklus (B).
Year 1 includes GP offices participating for the first time using semiquantitative instruments (n=635). Year 15 includes GP offices participating in 15 EQASs using semiquantitative instruments (n=10).

Table 2 Effect on analytical quality of changing to a different POC instrument.

	"Poor" trueness, n	"Poor" trueness the following year, n (%, 95% CI)
Total number of participants with "poor" trueness	199	49 (25)
Changing to a different POC instrument the following year	29	3 (10, 0–20)
Not changing to a different POC instrument the following year	170	46 (27, 21–33)
	"Very good" trueness, n	**"Very good" trueness the following year, n (%, 95% CI)**
Total number of participants with "very good" trueness	3784	3627 (96)
Changing to a different POC instrument the following year	151	138 (91, 88–94)
Not changing to a different POC instrument the following year	3633	3489 (96, 96–97)

percentage of GP offices never performing IQC decreased from 48% to 6% ($p<0.001$), and the percentage of participants using the DCA instruments increased from 24% to 42% ($p<0.001$), whereas the percentage using MICRAL-TEST decreased from 71% to 2% ($p<0.001$). The DCA line (including all DCA instruments on the market at the time of the study) performed better than each of the other quantitative instruments regarding both trueness and precision in more than 70% of the surveys (see Supplemental Data, Table 2). The profession of the person who performed the analysis, the frequency of IQC, the number of general practitioners at the GP office, and the number of analyses per month did not influence the participants' assessments.

Expiration date of the reagents

Participants using semiquantitative instruments used expired reagents significantly more often than those using quantitative instruments: 13% vs. 2%. Concerning quantitative instruments, GP offices using reagents with more than 3 months to expiration performed significantly better than those using reagents expiring within 3 months ($p<0.001$) or using reagents that had already expired ($p<0.001$). Similar findings were not seen for participants using semiquantitative instruments (results not shown).

Discussion

The value of participating in EQA programs has been emphasized by several organizations [15, 16], and in some countries, such participation is also mandatory [17]. Although participating in POC EQA programs is voluntary in Norway, 99.7% of Norwegian GP offices are enrolled in the program.

The effects of participating in EQA programs are difficult to quantify, and few studies have addressed the issue. However, one study found that the between-laboratory variation can decrease as the number of years participating in EQA programs increases [18], and another study found that circulating calibrators from EQA organizers can decrease the between-laboratory variation [19]. In the present study, we showed that the number of years a GP office had participated in Noklus, using quantitative instrument, was inversely related to the percentage of "poor" results. The same trend was not found for participants using semiquantitative instruments (Figures 1B and 2B).

Noklus offers not only an EQAS, but also an overall quality assurance system for laboratories in primary health care. In addition to the EQAS, this includes laboratory consultants visiting the GP offices providing courses and advice and distributing recommendations from the Scandinavian Evaluation of Laboratory Equipment for Primary Health Care [20]. With respect to the individual guidance performed by Noklus, the EQAS is the key for detecting which GP offices need additional help and education. That the improvement in analytical quality is due not only to the change in instrument is exemplified by the DCA line (DCA 2000, DCA 2000+, and DCA Vantage), whereby the trend to improve is present even when the same instrument brand is used. Upgrades of DCA instruments should not influence the improvement seen in the DCA line, as such upgrades mainly pertain to the instruments' software [21].

Participation in Noklus influences when a GP office chooses to change its instruments because laboratory consultants visit and offer advise about what instrument to buy. This advice is based on study of literature, SKUP evaluations [20], and performance in the EQA.

For quantitative instruments, analytical quality improved for participants who were initially assessed as "poor" and subsequently changed their instrument to a new version or to a different brand. Similar results have been observed when changing tests for infectious mononucleosis [22]. Participants that initially received an assessment of "very good" and then changed to a new instrument type did not improve their analytical quality. This suggests that participants performing well should usually not change the instrument they are using. It is a positive trend that most participants currently choose quantitative instruments for analyzing UA (see Supplemental Data, Figure 1). Early on, Noklus advised GP offices to use quantitative instruments and have now recommended that the Norwegian government should not reimburse for semiquantitative UA tests.

Although the performance characteristics of both quantitative and semiquantitative instruments have improved in recent years (Figures 1A and 2A), it is important to bear in mind that due to the wider quality specifications, the performance of semiquantitative instruments is not in any way comparable with that of quantitative instruments (see Supplemental Data, Table 1). This may explain why we did not find that participants using such instruments benefit from participating in an EQAS to the same extent as those using quantitative instruments.

The DCA instruments yielded especially good results in the present study: DCA instruments performed significantly better than the other instruments as shown by logistic regression analysis (see Supplemental Data, Table 2), which is in line with reports from other studies [23–25]. The present study also showed that use of IQC in GP offices has increased, which is encouraging because maintaining day-to-day consistency is important for obtaining reliable analytical results.

We found that participants using reagents expiring in more than 3 months performed significantly better than those using expired or nearly expired reagents. Similar findings have been reported for reagents for POC infectious mononucleosis [22]. The UA analysis is performed relatively seldom (47% of GP offices performed ≤10 analyses/week in Norway in 2012) and studies have found that GP offices do not screen for microvascular complications as often as desired [26–31]. This may cause manufacturers to extend the expiration date of the reagents. Other potential reasons like inappropriate storage or transport conditions may also contribute.

A limitation of the present study is that method-specific target values had to be used because the samples were kept at −80 °C a few days, or in a few cases up to 2 years, before the survey. Freezing could affect the UA methods in different ways [14], but studies indicate that the effects are probably not dependent upon the length of the freezing period [9, 32–35]. Thus, it is not possible to state unequivocally whether the differences between the mean values for the different methods are due to non-commutability or to "real" differences between the methods. However, there is currently no reference method or reference material for UA, making a "true" trueness evaluation impossible, although a candidate reference method and reference materials for UA has recently been developed [13]. Another limitation of the present study was that only UA and not ACR results were determined because only one instrument brand reported ACR throughout the entire study period. However, this is unlikely to have greatly altered the results because the assessments of UA and ACR were consistent in about 90% of the cases (results not shown). A major strength of the present study is that it includes data from a long period and from almost all Norwegian GP offices using the same type of EQA samples. Thus, all of the instruments measuring UA on the Norwegian market during that period were included, as were several changes in reagent lot numbers.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Financial support: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. American Diabetes Association. Standards of medical care in diabetes 2013. Diabetes Care 2013;36:S11–66.
2. International Diabetes Federation. Global guideline for type 2 diabetes. Available from: http://www.idf.org/webdata/docs/IDF GGT2D.pdf. Accessed 1 November, 2013.
3. National Kidney Foundation. KDOQI clinical practice guidelines and clinical practice recommendations for diabetes and chronic kidney disease. Am J Kidney Dis 2007;49:42–6.
4. Newman DJ, Mattock MB, Dawnay AB, Kerry S, McGuire A, Yaqoob M, et al. Systematic review on urine albumin testing for early detection of diabetic complications. Health Technol Assess 2005;9:iii–vi, xiii–163.

5. Viberti GC, Hill RD, Jarrett RJ, Argyropoulos A, Mahmud U, Keen H. Microalbuminuria as a predictor of clinical nephropathy in insulin-dependent diabetes mellitus. Lancet 1982;1:1430–2.
6. Arnlov J, Evans JC, Meigs JB, Wang TJ, Fox CS, Levy D, et al. Low-grade albuminuria and incidence of cardiovascular disease events in nonhypertensive and nondiabetic individuals: the Framingham Heart Study. Circulation 2005;112:969–75.
7. Xu J, Knowler WC, Devereux RB, Yeh J, Umans JG, Begum M, et al. Albuminuria within the "normal" range and risk of cardiovascular disease and death in American Indians: the Strong Heart Study. Am J Kidney Dis 2007;49:208–16.
8. Andrassy KM. Comments on 'KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease'. Kidney Int 2013;84:622–3.
9. Miller WG, Bruns DE, Hortin GL, Sandberg S, Aakre KM, McQueen MJ, et al. Current issues in measurement and reporting of urinary albumin excretion. Clin Chem 2009;55:24–38.
10. Libeer JC. Role of external quality assurance schemes in assessing and improving quality in medical laboratories. Clin Chim Acta 2001;309:173–7.
11. International Organization for Standardization. Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO 13528. Geneva, Switzerland: International Organization for Standardization, 2005.
12. Joint Committee for Guides in Metrology. Basic and general concepts and associated terms. International vocabulary of metrology (VIM) 2008. Available from: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf. Accessed November 1, 2013.
13. Lieske JC, Bondar O, Miller WG, Bachmann LM, Narva AS, Itoh Y, et al. A reference system for urinary albumin: current status. Clin Chem Lab Med 2013;51:981–9.
14. Schultz CJ, Dalton RN, Turner C, Neil HA, Dunger DB. Freezing method affects the concentration and variability of urine proteins and the interpretation of data on microalbuminuria. The Oxford Regional Prospective Study Group. Diabet Med 2000;17:7–14.
15. International Federation of Clinical Chemistry. Fundamentals for external quality assessment (EQA). Guidelines for improving analytical quality by establishing and managing EQA schemes. Examples from basic chemistry using limited resources. Available from: http://www.ifcc.org/ifccfiles/docs/Fundamentals-for-EQA.pdf. Accessed November 1, 2013.
16. CLSI. Using proficiency testing to improve the clinical laboratory: approved guideline, 2nd ed. CLSI Document GP27-A2. Wayne, PA: CLSI, 2007.
17. Buttner J, Borth R, Boutwell JH, Broughton PM, Bowyer RC. Approved recommendation (1983) on quality control in clinical chemistry. Part 3. External quality control. J Clin Chem Clin Biochem 1983;21:885–92.
18. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples – five year experience from the Netherlands. Clin Chim Acta 2012;414:234–40.
19. Cobbaert C, Weykamp C, Baadenhuijsen H, Kuypers A, Lindemans J, Jansen R. Selection, preparation, and characterization of commutable frozen human serum pools as potential secondary reference materials for lipid and apolipoprotein measurements: study within the framework of the Dutch project "Calibration 2000". Clin Chem 2002;48:1526–38.
20. SKUP. Scandinavian evaluation of laboratory equipment for primary health care. Available from: http://www.skup.nu/. Accessed November 1, 2013.
21. Szymezak J, Leroy N, Lavalard E, Gillery P. Evaluation of the DCA Vantage analyzer for HbA 1c assay. Clin Chem Lab Med 2008;46:1195–8.
22. Skurtveit KJ, Bjelkaroy WI, Christensen NG, Thue G, Sandberg S, Hoddevik G. [Quality of rapid tests for determination of infectious mononucleosis]. Tidsskr Nor Laegeforen 2001;121:1793–7.
23. Shephard MD, Barratt LJ, Simpson-Lyttle W. Is the Bayer DCA 2000 acceptable as a screening instrument for the early detection of renal disease? Ann Clin Biochem 1999;36:393–4.
24. Omoruyi FO, Mustafa GM, Okorodudu AO, Petersen JR. Evaluation of the performance of urine albumin, creatinine and albumin-creatinine ratio assay on two POCT analyzers relative to a central laboratory method. Clin Chim Acta 2012;413:625–9.
25. Shephard MD, Gill JP. An innovative Australian point-of-care model for urine albumin: creatinine ratio testing that supports diabetes management in indigenous medical services and has international application. Ann Clin Biochem 2005;42:208–15.
26. Harris SB, Stewart M, Brown JB, Wetmore S, Faulds C, Webster-Bogaert S, et al. Type 2 diabetes in family practice. Room for improvement. Can Fam Physician 2003;49:778–85.
27. Miller KL, Hirsch IB. Physicians' practices in screening for the development of diabetic nephropathy and the use of glycosylated hemoglobin levels. Diabetes Care 1994;17:1495–7.
28. Fagnani F, Souchet T, Labed D, Gaugris S, Hannedouche T, Grimaldi A. Management of hypertension and screening of renal complications by GPs in diabetic type 2 patients (France – 2001). Diabetes Metab 2003;29:58–64.
29. Mainous AG 3rd, Gill JM. The lack of screening for diabetic nephropathy: evidence from a privately insured population. Fam Med 2001;33:115–9.
30. Hueston WJ, Scibelli S, Mainous AG 3rd. Use of microalbuminuria testing in persons with type 2 diabetes: are the right patients being tested? J Fam Pract 2001;50:669–73.
31. Kraft SK, Lazaridis EN, Qiu C, Clark CM Jr. Marrero DG. Screening and treatment of diabetic nephropathy by primary care physicians. J Gen Intern Med 1999;14:88–97.
32. MacNeil ML, Mueller PW, Caudill SP, Steinberg KK. Considerations when measuring urinary albumin: precision, substances that may interfere, and conditions for sample storage. Clin Chem 1991;37:2120–3.
33. Brinkman JW, de Zeeuw D, Lambers Heerspink HJ, Gansevoort RT, Kema IP, de Jong PE, et al. Apparent loss of urinary albumin during long-term frozen storage: HPLC vs immunonephelometry. Clin Chem 2007;53:1520–6.
34. Graziani MS, Gambaro G, Mantovani L, Sorio A, Yabarek T, Abaterusso C, et al. Diagnostic accuracy of a reagent strip for assessing urinary albumin excretion in the general population. Nephrol Dial Transplant 2009;24:1490–4.
35. Hara F, Nakazato K, Shiba K, Shimoda J, Kojima T, Fukumura Y, et al. Studies of diabetic nephropathy. I. Effects of storage time and temperature on microalbuminuria. Biol Pharm Bull 1994;17:1241–5.

Supplemental Material: The online version of this article (DOI: 10.1515/cclm-2014-0483) offers supplementary material, available to authorized users.

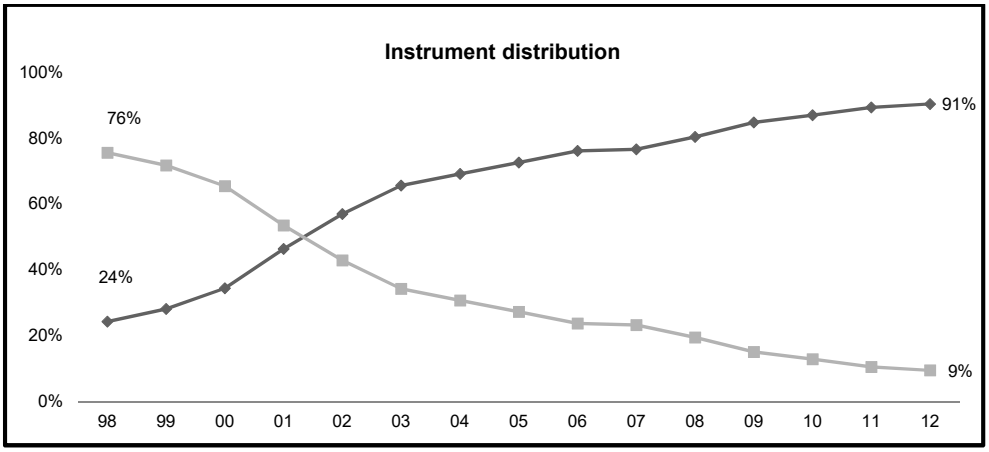
Supplemental file to the paper

Table 1. Noklus criteria for assessing trueness, precision and accuracy in the UA EQA.	
	Criterion
Quantitative instruments	
Target interval	Within ±2 mg/L of the method–specific target value
Trueness	
Very good	Mean value of duplicate measurements within ±7%
Acceptable	Mean value of duplicate measurements within target ±15%
Poor	Mean value of duplicate measurements outside ±15%
Precision	
Very good	Difference between duplicate measurements: ≤target value lower limit for "very good" trueness
Acceptable	Difference between duplicate measurements: values between "very good" and "poor" precision
Poor	Difference between duplicate measurements: target value lower limit for "acceptable" trueness
Semiquantitative instruments	
Accuracy	
Good	The categories with >30% of the answers
Questionable	The "neighboring" categories
Poor	The remaining categories
Quantitative instruments	**Example**
Target value	30 mg/L

Target interval	28–32 mg/L
Trueness	
Very good	26 (28 × 0.93) – 34.2 (32 × 1.07) mg/L
Acceptable	23.8 (28 × 0.85) –25.9 mg/L or 34.3 – 36.8 (32 × 1.15) mg/L
Poor	<23.8 (28 × 0.85) mg/L or >36.8 (32 × 1.15) mg/L
Precision	
Very good	Difference between duplicate measurements ≤4 (30– 26) mg/L
Acceptable	Difference between duplicate measurement range: 4.1– 6.2 mg/L
Poor	>6.2 (30–23.8) mg/L
Semiquantitative instruments (exemplified by Clinitek)	
Target value	30 mg/L (possible categories 10/30/80/150 mg/L)
Accuracy	
Good	The category “30 mg/L”. The interval ranging from 20 to 55mg/L
Questionable	The category “10 mg/L” and “80 mg/L”. The interval ranging from 0 to 19 mg/L and from 56 to 115 mg/L
Poor	The category 150 mg/L. Results >115 mg/L

Table 2. Results from logistic regression concerning trueness and precision*.				
			Trueness	Precision
POC instrument				
DCA all	Better than	NycoCard Reader II	9/9	9/9
DCA all	Better than	QuikRead all	5/7	5/7
DCA all	Better than	HemoCue Urine Albumin	6/8	6/8
DCA all	Better than	HemoCue 201	6/7	6/7
DCA all	Better than	Afinion	5/6	6/6
Afinion	Better than	HemoCue 201	3/6	

Number of EQA surveys with significantly better trueness and precision compared to number of EQA surveys where both instruments were included (number of participants >10). Only significant differences occurring in ≥50% of the surveys have been included. DCA all includes the DCA 2000, DCA 2000+ and DCA Vantage instruments. QuikRead all includes the QuikRead U-ALB and QuikRead101.



Supplemental file: Figure 1

II

III

Errata for
External quality assurance of point-of-care testing

Participant performance and commutability of the control material

Tone Bukve

Thesis for the Degree of Philosophiae Doctor (PhD)
at the University of Bergen

9/8-21 Tone Bukve (date sign. Candidate)

12/8-21 (date sign. Faculty)

Errata

Page 7: The text «besides the ones mentioned previously» has been replaced with «in addition to».

Page 10: The article «the» has been inserted prior to «MPs» and the word «additional» has been replaced with «few».

Page 13: The margins have been adjusted.

Page 19: The text «Table 2» and the phrase «Potential problems causing unacceptable EQA results» have been removed.

Page 26: The phrase «no POC MPs for use» has been replaced with «whereas none of the standards address POC testing performed». The word «actually» has been removed.

Page 31: The text «or not» has been replaced with «but».

Page 33: The word «this» has been replaced with «the».

Page 34: The text «that is» has been removed. The word «as» has been replaced with «by».

Page 36: The text «majority of» has been replaced with «most».

Page 39: The text «the manner in which» has been replaced with «how».

Page 40: The word «are» has been replaced with the text «must be taken into account. Examples of».

Page 45: The phrase «be able to» has been replaced with «can». The article «the» has been inserted prior to the word «participants».

Page 64: The text «Table 3» has been replaced with «Table 2».

Page 65: The size of the columns has been edited. The text «Table 3» has been replaced with «Table 2».

Page 66: The text «Table 4» has been replaced with «Table 3».

Page 75: The text «Figure 3 and Table 3» has been replaced with «Table 4».

Page 76: The text «Table 5» has been replaced with «Table 4».

Page 77: The text «to 2015» has been inserted. The phrase «At this time» has been changed to «In 2006».

Page 78: The phrase «As a consequence» has been replaced with «therefore».

Page 89: The phrase «As a result of» has been changed to «because of».

Page 95: The text «have to» has been replaced with «must».

Page 97: The phrase «the analytical quality of the participants» has been changed to «the participant`s analytical quality».

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230865811 (print)
9788230856338 (PDF)