papers at core.ac.uk

HUCH Laboratory Diagnostics
Hospital for Children and Adolescents
Helsinki University Central Hospital, Finland

Medix Laboratories Ltd.
Espoo, Finland

MEASUREMENT TOOLS FOR QUALITY ASSURANCE IN MEDICAL LABORATORIES

by

Solveig Linko

Academic dissertation

To be publicly discussed by permission of
the Medical Faculty of the University of Helsinki,
in the auditorium of Niilo Hallman,
Helsinki University Central Hospital,
Hospital for Children and Adolescents, Stenbäckinkatu 11, Helsinki
on February, 8th 2003, at 12 o'clock noon.

Helsinki 2003

Supervised by

Docent Jaakko-Juhani Himberg, MD, PhD
University of Helsinki
Faculty of Medicine
HUCH Laboratory Diagnostics
Helsinki University Central Hospital
Helsinki, Finland

Docent Eino Puhakainen, PhD
University of Helsinki
Department of Biosciences
HUCH Laboratory Diagnostics
Helsinki University Central Hospital
Helsinki, Finland

Reviewed by

Docent Hannu Jokela, PhD
University of Tampere
Department of Biosciences
Tampere, Finland

Docent Raija Puukka, PhD
University of Oulu
Department of Biochemistry
Oulu, Finland

Official opponent

Docent Erkki Seppälä, MD, PhD
University of Tampere
Faculty of Medicine
Tampere, Finland

ISBN 952-91-5530-1 (paperback)
Grafia Oy, Turku 2003

ISBN 952-10-0897-0 (PDF-version)
http://ethesis.helsinki.fi, Helsinki 2003

Per béllii ad astra

ACKNOWLEDGEMENTS

This study was carried out from 1998 to 2002 at Diacor, Deaconess Institute Clinical Laboratory in Helsinki, Medix Laboratories Ltd. in Espoo, Institute for Reference Materials and Measurements in Geel, and at HUCH Laboratory Diagnostics, (Hospital for Children and Adolescents, Helsinki University Central Hospital).

I highly appreciate the opportunity given to me at HUCH Laboratory Diagnostics, Helsinki University Central Hospital to finalize my work. My special thanks for support and encouragement are given to Professor Lasse Viinikka, M.D., the Head of HUCH Laboratory Diagnostics, to Professor Ulf-Håkan Stenman, M.D., and to Docent Martti Syrjälä, M.D., at HUCH Laboratory Diagnostics.

My supervisors Docent Jaakko-Juhani Himberg, M.D., Ph.D. and Docent Eino Puhakainen, Ph.D. are to thank for their precious time and good advice.

I wish to thank my co-workers Professor Linda Thienpont, Dr. Dietmar Stöckl, and Dr. Andre De Leenheer at the University of Ghent, my sister Docent Linnéa Linko, Ph.D., at the University of Turku, Docent Eero Taskinen, M.D., Dr. Päivi Kärkkäinen, M.D. and Professor Seppo Sarna, M.D., at the University of Helsinki.

I am inspired to express my sincere gratitude to Professor Veikko Komppa, at VTT Processes, Materials and Chemicals, who guided me to the footprints of metrology and has supported my work in such a way that cannot be measured.

Dr. Philip D.P. Taylor at the Institute for Reference Materials and Measurements,

Docent Ulf Örnemark, Ph.D. at the University of Upsala, and Dr. Rüdiger Kessel made me certain of one thing, which is most valuable: "We are still uncertain".

Docent Raija Puukka, Ph.D. and Docent Hannu Jokela, Ph.D. who reviewed my manuscript supplied with good points and constructive criticism. I am grateful for this advice.

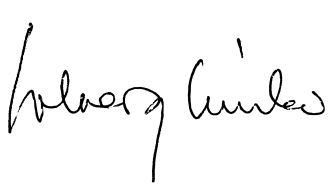
I want to thank Ms. Aila Koponen for her invaluable technical help.

I am lucky to have so many friends. Special thanks for technical assistance, in its many ways are given to Dr. Helene Markkanen, Med. Lic, Mrs. Liisa Pohjanen, Mrs. Marja Leena Laipio M. Sc., Mrs. Marita Junna, Ms. Siiri Elomaa, Mrs. Liisa Melamies, Phil.Lic, Paavo Tammisto, M.Sc., Mrs. Aino Salla. For the "material support" I wish to thank my husband Hannu Vartiala, Mr. Esa Passi, Mr. Seppo Räisänen (†), Mr. Seppo Tamminen, Mr. Sami Tuovinen, and Mr. Veikko Huhtaluoma. I wish to thank all the Medix Diacor people who have allowed this thesis to be undertaken.

This study was supported by grants from Biomedicum-Helsinki Foundation, Suomen Kemistiliitto, Association of Finnish Clinical Biochemists and funding from the European Commission.

Finally, I want to express my warmest feelings to my family for their tremendous support in such a many ways. Special thanks to Jaques le Roi.

Helsinki, December 2002



MEASUREMENT TOOLS FOR QUALITY ASSURANCE IN MEDICAL LABORATORIES

LIST OF THE ORIGINAL COMMUNICATIONS

I Linko S, Himberg J-J, Thienpont L, Stöckl D, De Leenheer A. Assessment of the state-of-the-art trueness and precision of serum total-calcium and glucose measurements in Finnish laboratories - the *QSL-Finland -study*. *Scand J Lab Invest* 1998; 58:229-240.

II Linko S, Örnemark U, Kessel R, Taylor PDP. Evaluation of uncertainty of measurement in routine clinical chemistry – applications to determinations of the substance concentration of calcium and glucose in serum. *Clin Chem Lab Med* 2002; 40(4):391-398.

III Linko S, Linko L, Himberg J-J. Self-monitored blood glucose – the need for quality goals and the role of clinical laboratory. Submitted for publication.

IV Linko S, Taskinen E, Sarna S, Kärkkäinen P. Factors affecting the cytology outcome of Pap smears - a brief approach to internal quality control in private cytopathology laboratory practice. *APMIS* 2001; 109:685-92.

V Linko S. Automated Ion Selective Measurement of Lithium in Serum. A Practical Approach to Result Level Verification in a Two-way Method Validation. *Accred Qual Assur* 2001; 6:31-36.

VI Linko S. Internal audits in private medical laboratory practice - a Finnish experience. *Accred Qual Assur* 2002; 7:55-59.

LIST OF ABBREVIATIONS AND ACRONYMS

1. Abbreviations for associations, committees, organisations and laboratories

ANSI	American National Standards Institute
BIPM	International Bureau for Weights and Measures
CAP	College of American Pathologists
C-AQ	Committee for Analytical Quality
CCHSA	Canadian Council on Health Services Accreditation
CCQM	Consultative Committee for Amount of Substance
CGPM	Conférence Générale des Poids et Mesures
CIPM	Comité International des Poids et Mesures
CITAC	Co-Operation on International Traceability in Analytical Chemistry
EA	European co-operation for Accreditation
EC	European Community
EGE-Lab	European Group for the Evaluation of Reagents and Analytical Systems in Laboratory Medicine
EQALM	European Committee for External Quality Assessment Programmes in Laboratory Medicine
EURACHEM	European Association for Analytical Chemistry
FDA	U.S. Food and Drug Administration
FINAS	Finnish Accreditation Service
GHTF	Global Harmonization Task Force
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IRMM	Institute for Reference Materials and Measurements
ISO	International Organization for Standardization
JCAHO	Joint Commission on Accreditation of Healthcare Organizations
LAP	Laboratory accreditation programme
NATA	National Association of Testing Authorities
NCCLS	National Committee for Clinical Laboratory Standards
NIST	National Institute of Standards and Technology (U.S.)
OECD	Organisation for Economic Co-operation and Development
SWEDAC	Swedish Accreditation
TAG	Technical advisory group
TC	Technical committee
WG	Working group

2. Acronyms

2.1 Acronyms used in equations and calculations

B_A	Analytical bias
C.I.	Confidence Interval
CV_A	Analytical coefficient of variation
CV_G	Inter-individual biological variation
CV_I	Intra-individual biological variation
CV_{PRE}	Pre-analytical variation
FN	False Negative

FP	False Positive
RCV	Reference Change Value
TE_a	Total allowable Error
TN	True Negative
TP	True Positive

2.2 Other acronyms

CLIA	Clinical Laboratory Improvement Amendments
EN	European Standard
EQA	External Quality Assessment
GLP	Good Laboratory Practice
GUM	Guide to the expression of Uncertainty in Measurement
ID-GC-MS	Isotope Dilution- Gas Chromatography- Mass Spectrometry
IMEP	International Measurement Evaluation Programme
IQC	Internal Quality Control
IVD	*In Vitro* Diagnostic (medical) Device
Pap	Papanicolaou (smear)
POCT	Point-Of-Care-Testing
PT	Proficiency Testing
SI	System International d`Unites
SMBG	Self Monitoring of Blood Glucose

ABSTRACT

Measurement quality in medical laboratories was studied with a set of six quality tools: **I** Use of reference methods and single donation sera, **II** Evaluation of uncertainty of measurement, **III** Assessment of performance characteristics, **IV** Internal quality control, **V** Method validation and **VI** Internal audits.

Using single donation whole blood and sera, routine patient serum and cervical smear samples, own observations from measurements and questionnaires, manufacturers' specifications, and data from literature carried out material sampling.

Common clinical chemistry routine methodology for serum total calcium, glucose and lithium, gynecological cytopathology for Papanicolaou tests and reference methodology for total calcium and glucose served as the basis for the methods of measurement used. The applied *in vitro* diagnostic medical devices consisted of system-dependent and system-independent calibrators with automated analyzing systems, patient-of-care testing meters intended for self-monitoring of glucose, control material for daily and proficiency testing purposes, and traceable reference materials.

Fit-for-purpose statistical methods and software were applied to classify the obtained data. The principles of international standards and guides were followed in all parts of the study. The present work was accomplished under co-operation of several European laboratories and institutes.

The use of the set of tools revealed common important characteristics and points of quality assurance in medical laboratories across the study. Reference methods and single donation sera were excellent tools for demonstrating laboratory performance in terms of

the state-of-the-art accuracy and trueness in Finnish laboratories. The importance of personnel skills, method validation and feedback meetings from internal quality control was emphasized. Need for training to perform internal audits was shown.

This study showed the necessity of the legislative control over the industry around *in vitro* diagnostic medical devices. Measurement quality is closely related to traceability, measurement hierarchy of metrology and evaluation of measurement uncertainty. The acceptable levels of uncertainty should be expressed as quality goals, which should be based on biological variation and medical needs.

Keywords: reference methods, quality, uncertainty, IQC, Pap smear, self-monitoring of blood glucose, SMBG, in vitro diagnostic medical devices, IVDs, lithium, validation, ISO/IEC 17025, internal audits

INTRODUCTION

The operation of laboratory medicine in patient care, monitoring and diagnosis is strongly linked to measurements and observations applied to laboratory samples. Producing reliable results within a reasonable turnaround time is the ultimate responsibility of medical laboratories. The total quality of the laboratory service, from the pre-analytical phase through the analytical phase to reporting, is to support the clinicians' decision-making.

The proper management of laboratory processes needs supervised personnel doing the right things in the right way. Well-defined rules are necessary for this management, as the numerous processes range from sample taking to reporting. International standards, guides and legislation support the establishment and implementation of quality systems. The quality of laboratory results, as being the end products of the process, thus strongly reflects the internal efficiency and the outcome of quality assurance.

Several hundreds of laboratory investigations exist in the production repertoire of modern medical laboratories. Indeed, the analysing techniques vary from quantitative high automation to qualitative manual observations, the latter being not of any less importance. Awareness and demonstration of the existing quality level of all results is of the utmost necessity. In the global harmonisation of medical laboratory results, demonstrating the traceability when possible is of great importance.

Quality is not a recent concept and medical laboratories have long traditions in demonstrating the reliability of the laboratory results. The primary objective of this study was to introduce a set of quality assurance tools and to demonstrate the appropriate use of

them in quality assurance. The set of tools reflects only some, of the crucial quality actions taken in medical laboratories.

In routine quantitative chemistry, the state-of-the-art accuracy of serum calcium and glucose analyses was shown in Finnish medical laboratories utilising reference methods and commutable control material (Tool **I**). In addition to accuracy, i.e. trueness and precision, the reliability of these common chemistry analyses was evaluated and expressed as the uncertainty of measurement utilising data and observations from laboratory routine measurements and data from manufacturer combined to a software application intended for this use (Tool **II**). Assessment of analytical performance characteristics is proposed in the example from Self-Monitoring-of-Blood-Glucose (SMBG) representing Point-Of-Care-Testing (POCT) (Tool **III**) and method validation from therapeutic drug monitoring (Tool **V**). When introducing internal quality control (IQC) to the set of tools, patient-related factors affecting re-screening of cytopathology samples and agreement in senior pathologists´ reviews was highlighted (Tool **IV**).

In the context of implemented and accredited quality systems, continuous quality improvement has to cover all quality processes including internal audits (Tool **VI**). Amending and upgrading this quality assurance tool has to derive from the opinions of the laboratory personnel.

REVIEW OF THE LITERATURE

1. Metrology in laboratory medicine

1.1 The infrastructure of metrology

Metrology is defined as `science of measurement´ including all aspects both theoretical and practical with reference to measurements, whatever their uncertainty, and in whatever fields of science of technology they occur[1].

In 1875, the Convention of the Metre (Convention du Métra) was signed by 17 nations in Paris as the necessity for global comparability was arisen[2,3]. Today, after slight modifications in 1921, this diplomatic treaty between fifty-one member states outlines an international measurement infrastructure with bodies/organizations and with links to national measurement institutes. The Convention gives authority to the Conférence Générale des Poids et Mesures (CGPM), the Comité International des Poids et Mesures (CIPM) and Bureau International des Poids et Mesures (BIPM) to act in matters of world metrology. Together with consultative committees the BIPM as being the international center for metrology, organizes the daily work. The consultative committee for amount of substance (CCQM) has existed since 1995. The demand for measurement standards of ever-increasing accuracy, range and diversity and the need to demonstrate equivalence between national measurement standards are of particular concern in this work. The Mutual Recognition Arrangement linked to national measurement institutes aims to increase the knowledge about the agreement between national measurement standards and specific measurements. Inter-laboratory comparisons or key comparisons coordinated by CCQM in the field of chemistry are important acts taken on this matter.

1.2 The International System of Units (SI)

In 1960, the System International d` Unites, SI became[4]. The seven dimensionally independent base units are the meter, the kilogram, the second, the ampere, the Kelvin, the mole, and the candela. Today, the kilogram, also known as `Le Grand K´, is the only remaining base SI unit defined by a man-made artifact. The global comparability of analytical results is in principal established by SI[5]. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) has recommended the use of SI in laboratory medicine due to the general acceptance that standardization of routine measurements should be done by agreement on common basis of metrology[6].

1.3 Traceability

Traceability is internationally defined[1] as `the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties´. This concept can be understood as a property of the value of the result of a measurement[5, 7]. Attaining traceability means providing measurement comparability, i.e. the ability to compare measurements on a global basis. It is stated that comparability is not only a problem of traceability to SI units or standards used. Moreover, the concept of comparability is connected to reference materials, method validation, and proficiency testing (PT)[8].

Traceability in laboratory medicine is not a new issue. In practice, traceability to SI means that measurement systems have to be designed in a way that they produce results closely related to the true value. A meaningful measurement system in five distinct parts has already been introduced in the 1970s[9, 10]:

Part 1. A rational, self-consistent system of units of measurement (e.g. the SI system)

Part 2. The material to realize in daily practice the defined units and their derivatives (e.g. the certified reference materials)

Part 3. The availability of accurate methods of measurements, analysis, or test, based on the well-characterized materials of part 2 (e.g. IFCC reference methods)

Part 4. Field or applied methods of measurement, analysis or test (i.e. the methods applied on a large scale in everyday work)

Part 5. A method whereby the long-term integrity of the measurement system is assured (e.g. inter-laboratory comparisons, proficiency testing).

Another approach for establishing traceability of the complete analytical procedure is described in the EURACHEM/CITAC Guide on Traceability in Chemical Measurements[11]. The essential activities in establishing traceability are:

- Specifying the measurand and the acceptable uncertainty
- Choosing a measurement procedure of estimating the value
- Demonstrating, through validation, that the measurement procedure includes all the "influence quantities" that significantly affect the result, or the value assigned to a standard
- Identifying the relative importance of each influence quantity
- Choosing and applying appropriate reference standards
- Estimating the uncertainty

Today, the directive for *in vitro* diagnostic medical devices (IVDs)[12] strongly supports traceability in the European community. The essential elements of a calibration hierarchy to support full calibration traceability to SI are identified in the draft international standards for IVD calibration traceability, ISO/CD 17511[13] and ISO/CD 18153[14].

1.4 Reference materials and reference methods

ISO defines the term `reference material´[15]: Material or substance one or more of whose property values are sufficiently homogenous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials. It is characteristic for primary reference materials that they are highly purified chemicals and that they can be directly weighed or measured to produce a solution whose concentration is exactly known[16].

The term `reference method´ is defined[15] as: Thoroughly investigated measurement procedure, clearly and exactly describing the necessary conditions and procedures, for the measurement of one or more property values that has been shown to have trueness of measurement and precision of measurements in accordance with its intended use and that can therefore be used to assess accuracy of other measurement procedures for the same properties, particularly in permitting the characterization of a reference material.

The purpose of a measurement is to describe a property of the investigated material[17]. In a correct metrology measurement system, (i.e. measurement apparatus, reagents, and the calibrator), this is done by a defined measurement procedure using a calibrator, with an assigned and traceable value, anchoring the signal(s) through a measurement function to end up with a measurement result.

The causes of poor comparability of medical laboratory results are[18]:

- unspecific measurement procedures
- incorrect calibration of the measurement procedures
- inadequate definition of the quantities.

The incorrect calibration can be avoided by utilizing suitable (certified) reference materials and reference methods.

Aiming at the improvement of the result comparability, the awareness for the need of certified reference materials increased dramatically in the 1990s within many fields of analytical chemistry including the medical laboratory sector[19]. The Institute for Reference Materials and Measurements (IRMM) offers invaluable metrology support to the clinical chemistry sector[20] as well as to other analytical chemistry fields. Many certified reference materials for use at the medical laboratories are available[21, 22]. When reference materials are concerned, `commutability´ means the ability of the material to show inter-assay changes comparable to those observed in the measurement of the same analyte in human serum[23]. Therefore, a possible lack of commutability makes the certified reference materials useless[24].

The hierarchy of measurement methods[25] with SI units at the top level and with increasing bias is: definitive methods (bias ca. 0.1 to 1%), reference methods (bias ca. 1-3%), and routine methods (bias ca. 5-10%) (Figure 1). The CCQM defines `a definitive method´[26] as: A primary method of measurement is a method having the highest metrological qualities, whose operation can be completely described and understood, for which a complete uncertainty statement can be written down in terms of SI units. In a practical meaning, the measurement structure has been described by Tietz[10]. Definitive methods are those of highest quality used for validating reference methods and primary reference materials, i.e. reference materials of highest quality. The observed value obtained by the field method is linked to the true value obtained by the definitive method through the traceability chain[27, 28].

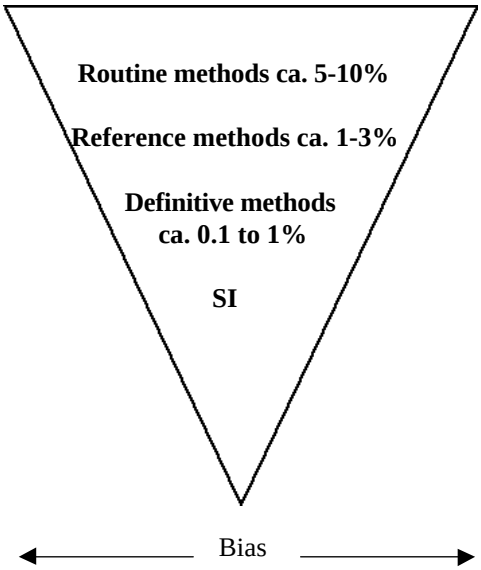


Figure 1. The three-level hierarchy of measurement methods described by Uriano and Cali[25].

Isotope dilution-mass spectrometry (ID-MS) is still the only adequate technique for the development of definitive methods. Applications of ID-MS have been widely used in clinical chemistry since the 1970s[29,30,31]. Some of the advantages of ID-MS are high precision (imprecision $\leq$ 0.2%), unbiased nature of the determination, high sensitivity (up to 10^{-12}g depending upon element and instrumentation), and high selectivity, i.e. possessing very few interferences[32]. In addition to the primary methods of measurements (i.e. ID-MS, gravimetry, titrimetry, coulometry, freezing point depression) many reference measurement procedures and candidate reference methods have been developed for clinical chemistry quantities[33, 34, 35, 36]. In 1998, European Committee for standardization defined the presentation of reference measurement procedures[37].

In conventional clinical chemistry serum, plasma and whole blood components have been traditionally the most frequently investigated materials. Most of the measurements still take place in laboratory environment despite the constant increase in near-patient testing[38]

and fast growth of POCT test systems in self-monitoring of diabetes, where no common basis for comparing the accuracy and precision of these instruments exist[31].

1.5 Measurements based on identification

Metrology and traceability inhere most often to quantitative analyses. Truly, the quantitative analysis performed deals with only one part of the measurement spectra concerned. According to a recent report, the IVD field routinely performs 400 to 600 different amounts of substances (analytes) with full calibration systems with traceability to SI currently existing for less than 30 (i.e. ca. 5%) of these analytes[39].

Pattern recognition, identification, subjective interpretation and classification are common practice in e.g. clinical pathology, microbiology and molecular biology, or forensic science. In forensic science, a set of principle of good practice in qualitative analysis has been reported[40]. Identification may be described as classification according to specific criteria[41]. It often pertains to subjective interpretation on limits such as satisfactory or unsatisfactory, above or below, or classification into ranges such as amount of particles or color intensity. In gynecological cytology, the Bethesda System[42] updated in 2001 outlines the terminology for reporting results of cervical cytology. This is crucial to harmonize and promote effective communication of cervical cytology results from the laboratory to clinicians and to avoid misunderstanding where `measurement´ and reporting is metrologically impossible.

2. The concept of quality and related issues

2.1 The definition of quality

The concept of quality can be described in numerous ways. According to the international standard ISO 8402, `quality´ is a totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs[43]. The achievement of satisfactory quality involves all stages of the quality loop as a whole. It is notified in the standard[43] that the needs may include: aspects of performance, usability, dependability (availability, reliability, maintainability), safety, environment, economics and aesthetics. The British Standard[44] points out the various meanings of the quality concept as it is used in different settings. Quality can be discussed: in a comparative sense, whenever products or services are ranked on a relative basis; in a quantitative sense as used in technical evaluations, or in the fitness-for-purpose sense, whenever products or services are evaluated in terms of their ability to satisfy a given need[45].

2.2 Development of quality thinking

The revolution of quality evolved as a consequence of World War II. The first applications of statistical quality control were taken into use in order to remove inadequate products intended for military purposes in the US. It became clear that high quality was produced as a result of inspection and testing procedures[46].

The history of quality thinking in health care can be traced back to the middle of 1800th century. The first colonial hospital, Pennsylvania Hospital, routinely tabulated the clinical outcomes of its patients in terms of symptoms, cure, or death[47]. At Massachusetts General Hospital, one further step was taken in the early 1900´s, as a surgeon, Ernest Codman developed the `end result system´. He identified reasons for less than optimal outcomes by categories such as `errors due to lack of technical knowledge or skill´, `lack of surgical

judgement´, `lack of care or equipment´, or `lack of diagnostic skills´[47]. In 1917, the American College of Surgeons took the initial attempt towards improved health care quality. This college established a hospital standardization programme[45]. The Minimum Standard was published in 1919 because of this work, which included specific requirements for diagnostic and therapeutic facilities. It states: The diagnostic and therapeutic facilities under competent supervision available for the study, diagnosis, and treatment of patients, should include, at least: (a) a clinical laboratory providing chemical, bacteriological, serological, and pathological services; (b) an X-ray department providing radiographic and fluoroscopic services.

During the pre-penicillin era in the 1930s, it became obvious to require external quality assessment (EQA) schemes for syphilis serology to limit the administration of arsenic and mercury and simultaneously to minimize toxic chemotherapy[48]. In 1950, Levey and Jennings introduced the use of control charts in medical laboratories[49] according to the ones previously used in industrial processes and known as Shewhart plots[50]. Westgard rules were developed in the early 1980s to facilitate the IQC run by the medical laboratory professionals[51]. Since the "quality-thinking" has grown in the fields of laboratory medicine as well as in the industry manufacturing diagnostic products[52]. Due to the positive attitude to the concept of quality, monitoring and assessing the laboratory performance has been run for decades by professionals in medical laboratory sector mainly by focusing on well-established quality tools[53], such IQC[51] and EQA[54,55].

2.3 Quality management

`Quality management is defined as the part of the overall management functions that determines and implements quality policy[43]. It was gradually realized that by doing the right things from the beginning, i.e. relevantly, timely, and effectively from economic point of view, would satisfy the needs set and stated by the many stakeholders of medical laboratories. This led to understanding of total quality management. Westgard's and his colleagues' work is a remarkable milestone in the development of total quality management in laboratory medicine[56].

Throughout the 1980s and the 1990s, much attention has been paid to principles of quality management associated with organizational structures, responsibilities, procedures, processes, and resources[57]. Two distinct systems of total quality management applicable to medical laboratories exist: accreditation and certification. By definition, `accreditation´ is a procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks[58]. On the contrary, `certification´ is a procedure by which a third party gives written assurance that a product, process or service conforms to a specific requirement[58]. In common usage to `accredit´ means to certify or guarantee someone or something as meeting required standards and to `certify` means to endorse or guarantee that certain required standards have been met[59]. The requirements are written in documents called 'standards` usually implemented by international or national organizations. The concept `standard´ is described as a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context[60].

2.4 Laboratory accreditation and certification

Globally, the organizations accrediting or certifying medical laboratories are of different types, i.e. governmental or authoritative organizations. The development in laboratory accreditation started, as it became clear to the United States Congress that unsatisfactory testing was performed within health care sector[61]. Consequently, the College of American Pathologists (CAP) initiated the first accreditation scheme in 1961 specially designed for medical laboratories: the Laboratory Accreditation Programme (LAP). Today, the CAP programme[62] is recognized by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO)[63] and has a decision authority under the Clinical Laboratory Improvement Amendments of 1988, CLIA´88[64].

In Australia the National Association of Testing Authorities (NATA) as a principal inspection agency, has experience with accreditation for over 50 years mainly for the benefit of Australian industry, government, and the community[65]. A medical testing program was established by NATA in 1983 to accredit pathology facilities. The Australian principles of accreditation follow the ISO 9000 standard series[66] and ISO/IEC Guide 25[67]. The Canadian Council on Health Services Accreditation (CCHSA) introduced a Client Centered Accreditation Program in 1995[68] focusing on the implementation of Total Quality Management in medical laboratories, but no federal approach has yet been developed.

In addition to patient care the scope of medical laboratory analyses are also included in medical trials. In this context, medical laboratories have to follow good laboratory practice (GLP) standards as tests are performed in the pre-clinical phase. In European countries the GLP directive is based on the principles based on the Organisation for Economic Co-operation and Development, the OECD guideline[69], while the laboratories

involved with medical trials are controlled by the U.S. Food and Drug Administration (FDA). The National Committee for Clinical Laboratory Standards (NCCLS) in the US contributes to guidelines for health care professionals and manufacturers in terms of GLP and medical laboratory testing[45].

To demonstrate the required quality procedures, European medical laboratories started to take actions during the 1990s in developing their quality systems according to EN 45001[70] based on the ISO/IEC Guide 25[67] or ISO 9000 standard series[66]. The first medical laboratories were accredited in Sweden in 1992 by SWEDAC, the Swedish accreditation body[71]. Since then, the number of accredited laboratories, representing disciplines of clinical chemistry, clinical microbiology, blood banking, and pathology has been growing exponentially in the Nordic countries. Today, more than twenty medical laboratories in Finland have fulfilled the accreditation requirements assessed by FINAS[72], the Finnish Accreditation Service that together with other accreditation bodies is a member of European co-operation for Accreditation (EA)[73]. In the United Kingdom, medical laboratories follow the national standards set by the Clinical Pathology Accreditation (CPA)[74], which serves also as the national accreditation body. Guides and recommendations were established in many countries by international, national, organizational, and professional groups to facilitate this demanding work[75, 76, 77, 78, 79, 80, 81, 82]. In addition to analytical issues, guidance for documenting and implementing some special actions has been taken in account. These actions include e.g. internal audits[76, 84], an important management tool which medical laboratories might not have been so familiar with before[83].

The new international standard ISO/IEC 17025, General requirements for testing and calibration laboratories[84], replaced the criteria of the EN 45001 and ISO/IEC Guide 25 standards for laboratory accreditation by the end of 2002.

In co-operation between ISO, the US standardization body, American National Standards Institute (ANSI), and the NCCLS, the ISO Technical Committee 212, ISO/TC212 has worked out the first International Standard for Quality management in the medical laboratories EN/ISO 15189[85]. The proposed standard has been prepared to specify the requirements for the quality management of a medical laboratory and to cover all examinations and provide guidance for laboratory procedures to ensure quality in medical laboratory examinations[86, 87]. It has been claimed that this standard will bring the quality management in medical laboratories closer to total quality management than previous standards[88].

2.5 Requirements set for manufacturers and products

It is declared in the Essential requirements, Annex 1 in the European Community, EC directive set for IVDs[12] that:

The devices must be designated and manufactured in such a way that they are suitable for the purposes referred to in Article 1(2)(b), as specified by the manufacturer, taking account of the generally acknowledged state of the art. They must achieve the performances, in particular, where appropriate, in terms of analytical sensitivity, diagnostic sensitivity, analytical specificity, diagnostic specificity, accuracy, repeatability, reproducibility, including control of known relevant interference, and limits of detection, stated by the manufacturer. The traceability of values assigned to calibrators and/or control values must be assured through available reference measurement procedures and/or available reference materials of a higher order.

The objective of IVDs´ design is to produce medically useful results. In this context, well-defined laboratory quality specifications play an important role. Manufacturers

conforming to ISO quality system standards must follow a disciplined design control when developing a new IVD system[89, 90, 91].

Design control has five general steps[92]:

- define user requirements
- translate user requirements into design specifications
- design and develop the product to satisfy the requirements and design specifications
- verify the product meets the specifications
- validate the product by demonstrating that the user requirements are met.

The prime objective of design control in the development of medical devices is to deliver the product to market economically and have it perform safely and effectively for its intended use[93]. There are close relations between industry and the users of IVDs, i.e. professionals at the laboratory workbench or patients performing self-monitoring[94].

The FDA Design control guidance[95] for medical device manufacturers has been developed by the contribution of the Global Harmonization Task Force[96] (GHTF) Study Group 3. The GHTF was formed in 1992 to further this effort. The GHTF includes representatives of the Canadian Ministry of Health and Welfare; the Japanese Ministry of Health and Welfare; FDA; industry members from the European Union, Australia, Canada, Japan, and the United States; and a few delegates from observing countries. The ultimate purpose of GHTF is to respond to the growing need for international harmonization in the regulation of medical devices. The CLIA'88 regulations categorize tests on the basis of the complexity of the test methodology[97]:

- waived tests
- tests of moderate complexity

- tests of high complexity.

3. Evaluation and expression of measurement quality

3.1 Method validation

According to the standardized definition[43] used for `validation´, this evaluation process is confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled. A process very close to validation, i.e. `verification´[43] is performed when a laboratory wishes to confirm that specified requirements have been fulfilled. In laboratory medicine, validation can be understood as an adequate examination of a laboratory or a POCT method of measurement intended for a clinical investigation, i.e. monitoring or diagnosis. Clinical laboratory professionals meet the need for selection and evaluation of either new or modified methods recurring. At the time, both standardized and non-standardized methods shall be covered[98, 99]. Good laboratory practice postulates well-established processes prior to method adoption to routine use. According to the modern approach of a new method introduction begins with establishment of need, method selection, and quality goal setting[27]. The six Valid Analytical Method -principles have been introduced in the EURACHEM Guide for The Fitness for Purpose of Analytical Methods[99]. The first principle stresses that analytical measurements should be made to satisfy an agreed requirement regarding measurements made under well-defined quality control and quality assurance procedures. Thus, an operational definition is needed first to agree on[46, 75].

Due to its demanding nature, the outlines of validation (and verification) shall consist of:

- planning, timing and follow-up
- performance according to reasonable schemes
- documentation

- reporting
- acceptance.

In the laboratory medicine field, the first method evaluation schemes[100] were introduced in the 1970s. Several experts and expert groups have since then worked out evaluation protocols for medical laboratories and IVD manufacturers[101, 102, 103, 104, 105].

3.2 Analytical performance

From the medical laboratory perspective, information on the reliability of results is necessary for several reasons. First, a laboratory professional has to evaluate the fulfillment of quality goals in method validation, establishing of IQC or in running daily quality control[106, 107, 108]. Secondly, it is important that the result of a measurement is accompanied with information of the error or uncertainty (within a defined confidence interval)[109, 110, 111]. Thirdly, the competence of the laboratory may be, and is often judged against the analytical performance in EQA[112] or third party assessment according to available international standards[84, 85]. Common understanding and expression of terms is important in any field of science and technology[113]. The pivotal ISO definitions characterizing analytical performance exist as the following concepts[114]:

- trueness - The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value
- precision - The closeness of agreement between independent test results obtained under stipulated conditions
- accuracy - The closeness of agreement between a test result and the accepted reference value

- uncertainty - An estimate attached to a test result, which characterizes the range of values within the true value, is asserted to lie.

All testing from pre-analytic phase to the reporting involves with error and uncertainty sources. In quantitative analyses, the reliability of the measurement quality is expressed as random error (i.e. precision) and systematic error (i.e. trueness, or bias)[109]. The combination of these two errors is comprehended as total error, TE (i.e. accuracy). For this reason, it is important to distinguish the difference between error and uncertainty as stressed in the available guides[115, 116, 117, 118]. Current international standards[84, 85] applied to medical laboratory accreditation and quality management describe clearly the requirements for evaluation and calculation of the uncertainty of measurement whenever possible. In modern laboratory practice, the expression of the uncertainty of measurement has become an inevitable concept[109, 119]. In qualitative analyses, the results are reported on a nominal or ordinal scale. Reports in laboratory medicine often include various categorical statements. Uncertainties of non-quantitative tests in many areas[41, 120, 121] are expressed as alternative reliability measures such as,

- false positive rate - $FP^{i}/(TN^{ii} + FP)$
- false negative rate - $FN^{iii}/(TPi^{v} + FN)$
- sensitivity - TP/(TP + FN)
- specificity - TN/(TN + FP)
- efficiency - (TP + TN)/(TP + TN + FP + FN)
- Youden index[122] - sensitivity + specificity - 100
- likelihood ratio - (1 - false neg. rate)/(false pos. rate)
- Bayes posterior probability[123],

[i]FP=false positive, [ii]TN=true negative, [iii]FN= false negative, [iv]TP=true positive.

3.3 Laboratory performance and quality goals

Medical laboratories have long traditions in demonstrating their analytical quality by means of EQA or PT[124]. The term EQA is more established in Europe among medical laboratory professionals. Since the early 1950s, EQA[125] has provided an essential quality assurance tool complementary to IQC in comparing the performance between laboratories. Starting with national EQA surveys the history of Labquality in Finland can be traced back to the early seventies[126]. ISO has published ISO/IEC Guide 43 on Proficiency Testing by Inter-laboratory Comparisons including examples of statistical methods for treatment of PT data[127].

Today, EQA schemes cover several disciplines in the medical laboratory sector worldwide. The goals of EQA are explicit[128, 129, 130, 131, 154]:

- to maintain the long-term accuracy of the analytical methods
- to evaluate participant performance
- to train participants.

In the European Community, the International Evaluation Programme (IMEP) organized by IRMM has promoted inter-laboratory comparisons in co-operation with reference/national metrology institutes and EQA organizers[132, 133, 134, 135].

Assessment of laboratory performance is closely linked to the set goals. A working group under the European Group for the Evaluation of Reagents and Analytical Systems in Laboratory Medicine (EGE-Lab) has recommended applying the biological approach as the basis for analytical specifications for routine method bias and precision, and total error[136]. The members of the External Quality Assessment Working Group A on analytical

quality goals in laboratory medicine recommends that the total allowable error, TE_a is calculated from the following equation[137]:

$TE_a = 1.65 * CV_A + [B_A]$ ($a<0.05$),

CV_A, and B_A are desirable quality specifications for imprecision[146] and bias[152].

Optimally, if the laboratory uses either two different methods (or two different instruments) for the same analyte, the allowable difference between them should not exceed one third of the intra-individual biological variation of the analyte[138].

In external quality assessment schemes (EQAS), biological data can be used to set specifications for the fixed limits of acceptance[139]:

EQA-limit $= 2.33 * CV_A + [B_A]$ ($a<0.01$)

However, the limits of acceptance and the criteria for setting the limits vary between countries[140, 141]. Further, it has been proposed that more stringent quality specifications, $TE_a \leq 1/5 *$ EQA-limit should be used when reference methods are concerned[142].

As monitoring of a patient is concerned, it is most important that the analytical variation[143], CV_A should not increase the variability of test results also influenced by the intra-individual biological variation, CV_I[144] and pre-analytical variation, CV_{PRE}[145].

The first concept for tolerable analytical variation was proposed in 1970[146]:

$CV_A < 0.5 * CV_I$.

In therapeutic drug monitoring, a model based on pharmacokinetic theory is used[147]:

$CV_A \leq 0.25[(2^{T/t} - 1)/(2^{T/t} + 1)] * 100$,

T is the time interval between doses and t is the average elimination half-life. The equation above shows that drugs with small dosing interval or long half-life require better

precision. The goal is that methods should have no bias and only true values should be generated. Therefore, goals for precision should be more or less equal to goals for total analytical error, desirably with index of fiduciality, i.e. $CV_A/0.5 * CV_I$, less than 1.0[148].

Reference change value, RCV must exceed the inherent variation due to biological, and analytical variation in the assessment of the patient's status[149]:

$RCV > 2^{0.5} \cdot z \cdot [CV_A^2 + CV_I^2]^{0.5}$,

z is for example, 1.96 for $p < 0.05$ and 2.56 for $p < 0.01$.

A clinician usually compares the test results either to clinical consensus guidelines, e.g. glucose[150], or to population-based reference limits[151] to be able to reveal illness as early as possible. The following definition for analytical bias, $[B_A]$ based on the group biological variation has been introduced[152]:

$[B_A] < 0.25 [CV_I^2 + CV_G^2]^{0.5}$,

CV_G is the inter-individual biological variation (i.e. between-subject variation).

This criterion for bias is interpreted as follows: 120 individuals are selected for a reference population as recommended by IFCC[153]. Then the maximum bias allowable to achieve the maximum acceptable percentage of the population outside each limit for the 0.90 confidence interval of each of the reference limits (mean +1.96 *s*), which equals to 4.4%.

It has to be emphasized that these currently available quality specifications are applied only for routine clinical chemistry and that no global specifications are available for POCT until now[154], despite of the on-going standardization work on establishing performance criteria for *in vitro* blood glucose monitoring systems[155]. Working Group 3 (WG 3) of the International Standards Organization Technical Advisory Group 212 (ISO

TAG 212) administered by NCCLS has developed a draft document on analytical goals[156]. According to this document, a well-established quality management plan must take into account economic and regulatory needs in addition to the most important basis of desirable analytical performance goals, the medical needs. A hierarchical approach to classification of strategies is presented in the Consensus Statement[157] from the Stockholm Consensus Conference in 1999. The objective of the Stockholm meeting was to reach consensus on the setting of global quality specifications in laboratory medicine. This was achieved successfully, resulting in unanimous agreement between the participants. Where available, and when appropriate for the intended purpose, models higher in the hierarchy are to be preferred to those at lower levels. The concept of such a hierarchy is described[158]. This hierarchy has also been proposed by the ISO/TC 212 WG 3 subgroup on `Analytical Performance Goals Based on Medical Needs´.

Quality goal setting in measurements based on pattern recognition and subjective interpretation, like cytopathology is more complex. First, the laboratory report is a result of several affecting factors and evaluated parameters. Contradictory reports appear on sensitivity and specificity of i.e. conventional Papanicolaou (Pap) testing by cytotechnologists[120, 159]. Self-evidently, the most stringent goal for pathology is that no false diagnosis should ever be done.

AIMS OF THE STUDY

The objective of this study was to identify and assess an adequate set of quality assurance tools in demonstrating and improving the reliability of measurements performed at modern medical laboratories in context with current international standards, guides and recommendations.

The set of tools and specific examples of their use were in the studies **I-VI**:

I **Tool:** Use of reference methods and single donation sera

Example: The use of this tool was demonstrated by assessing the state-of-the-art trueness and precision of serum total-calcium and glucose routine measurements in Finnish medical laboratories. The results were compared to reference method values obtained by ion chromatography and isotope dilution-gas chromatography-mass spectrometry (ID-GC-MS).

II **Tool:** Evaluation of the uncertainty of measurement

Example: The uncertainty of measurement was evaluated in the determination of the substance concentrations of calcium and glucose in serum in common routine clinical chemistry to emphasize the need for expressing measurement reliability.

III **Tool:** Assessment of performance characteristics

Example: The analytical performance of two different POCT meters commonly applied in SMBG was assessed. The performance characteristics were

compared to the available analytical goals and to the fit-for-purpose specifications.

IV **Tool:** Internal quality control, IQC

Example: Factors affecting the outcome of IQC procedures applied in cytopathology by pattern recognition and subjective interpretation as methods of measurement were investigated.

V **Tool:** Method validation and result level verification

Example: A method validation procedure was introduced in the field of therapeutic drug monitoring. The determination of the substance concentration of lithium in serum was taken as an example.

VI **Tool:** Internal audits

Example: Opinions of personnel on the internal audit process were surveyed at a medical laboratory.

MATERIALS AND METHODS

1. Material sampling

1.1 Serum samples

Single donation sera were obtained from six voluntary males (**I**). All donors were tested to be negative against human immunodeficiency 1 and 2 virus and Hepatitis C virus antibodies, Hepatitis B surface antigen and syphilis prior to any further actions taken. The blood donations took place under ethical conditions at the Helsinki University Central Hospital. After separating from the blood cells, the native serum was filtrated through a tuft of glass wool. The material obtained was carefully mixed, then divided into aliquots before freezing in tightly capped plastic tubes. This was to ensure the homogeneity and the stability of the analytes to be studied: calcium and glucose. Six aliquots of each sample, frozen on solid carbon dioxide, CO_2 were shipped to the reference laboratories at the University of Gent, Belgium and to the participating laboratories of QSL-Finland study (**I**).

Four voluntary healthy adults donated whole blood 3 x 5 ml (**III**). Two adults were fasting before phlebotomy and two were on normal diet. Four glucose levels were prepared: lower hypoglycemic, hypoglycemic, euglycemic and hyperglycemic level. The hypoglycemic levels derived from incubating the whole blood samples maximally 18 hours at room temperature. The euglycemic level was from one of the non-fasting donors and the sample was appropriate as such. The hyperglycemic level was obtained by spiking the blood drawn from the other non-fasting donor with 750 mmol/l of *D*-glucose. After level adjustments, all samples were treated in an equal way (**III**).

The Medix Laboratories Ltd. (Espoo, Finland) (Medix) patient sample pool served as the source of testing material for 62 serum lithium samples (**IV**). The reported serum lithium concentrations in mmol/l were recorded for method comparison (**V**).

During 1996-1999 the cytopathology sample pool at Medix was used in the sampling of 119 of 87409 Pap smears retrospectively double-screened by cytotechnologists and 354 of 87409 Pap smears reviewed by pathologists as internal quality control (**IV**). From the selected and double-screened Pap smears, the use of intra-uterine device, patient age ($\leq$ 47 year and > 47 year), and hormone replacement treatment were filed (**IV**).

1.2. Other test material

The raw absorbance data was obtained from two common routine clinical chemistry spectrophotometric methods used at Medix for determination of the amount of substance of calcium and glucose in serum: *o*-cresolphthalein complexone for calcium, and the enzymatic reference method with hexokinase for glucose (**II**). The necessary information about calibrators and instrument specifications were from the manufacturer of the measurement system (Roche Diagnostics Ltd., Mannheim, Germany) (**II**).

Laboratory personnel experiences from vertical and horizontal audits during 1996-2000 at Medix served as the basis of studying the fulfillment of common quality management procedures. The brief questionnaire resulted in 74 replies from 120 employees at Medix (**VI**).

2. Principles of measurements in routine methods

The following principles of measurement for the determination of calcium in the serum samples were atomic absorption spectrometry, liquid or reflectance absorption spectrophotometry, flame photometry, and potentiometry, ion selective electrode (**I, II**). Medix and other Finnish collaborators applied amperometric, absorption spectrophotometric and reflectance methods for serum and whole blood glucose measurements (**I, II, III**). Morphological investigation was the principle of measurement technique in the cytopathology study (**IV**).

Direct ion selective electrode applications were used for serum lithium measurement (**V**).

3. Reference methods

A primary reference method, ID-GC-MS was used for serum glucose (**I**). Ion chromatography was used to obtain the reference method values for serum calcium (**I**). The reference method for the investigation of Pap smears: Papanicolaou´s staining and cell morphology based on the Bethesda System was used routinely in the cytology study (**IV**). Reference method values for serum lithium PT samples were obtained operating under the principles of flame emission photometry (**V**).

4. *In Vitro* Diagnostic Medical Devices

4.1 Calibration devices

Both system-dependent and system-independent calibrators were used in the QSL-Finland study (**I**). The applied glucose and calcium methods were calibrated against commercially available *D*-glucose materials and respectively against calcium materials (**I**). Calibrator for automated systems (C.f.a.s.) (Roche Diagnostics Ltd.) was used for

calibration of the measurement of serum glucose hexokinase method (**II, III**) and respectively for the calibration of the *o*-cresolphthalein complexone method to measure calcium in serum (**II**).

The ion selective electrode setups: Cobas® Intergra ISE Module (Roche Diagnostics Ltd.), Microlyte 6 ISE (Kone Instruments, Espoo Finland) and Chiron 654 $Na^+/K^+/Li^+$ (Chiron Diagnostics Ltd., Halstead, Essex, U.K.) were equipped with their own system solutions and calibrators and were purchased from the respective manufacturers (**V**).

According to the manufacturer, GlucoTouch™ (Life Scan Inc., Milpitas, CA, U.S.A.) was "factory-calibrated" to the plasma glucose level with 21 calibration events against YSI 2700 Glucose Analyzer (YSI Incorporated, Yellow Springs, OH, U.S.A) (**III**). GA-1140 Glucose AUTO & STAT (KDK Corporation, Kyoto, Japan) was used as the reference system in the calibration of Super Glucogard II™ test strips (**III**).

4.2 Measurement devices

Medical laboratory instrumentation intended for routine use was applied in QSL-Finland study (**I**). Cobas® Integra 700 clinical chemistry analyzer (Roche Diagnostics Ltd.) equipped with ion selective electrode module was used for the measurements of serum calcium (**II**), serum lithium (**V**), and serum glucose (**II, III**). POCT glucose meters used for SMBG were GlucoTouch™ and Super GlucoGard II™ (**III**). The two other ion selective electrode setups were Microlyte 6 ISE and Chiron 654 $Na^+/K^+/Li^+$ (**V**). Common light microscopes represented measurement devices in the cytopathology study (**IV**).

4.3 Control material

PT samples of past schemes were purchased from Labquality Ltd. (Helsinki, Finland) and from Murex Biotechnology Ltd. (Dartford, Kent, U.K.), currently under the company name of Bio-Rad Laboratories Inc. (Hercules, CA, U.S.A). The PT samples were either liquid or lyophilized material and of human origin. Daytrol™ from Labquality Ltd. and system-dependent control materials from Roche Diagnostics Ltd. were used in daily IQC (**II, III, V**).

System-dependent control solutions were used in POCT: a one-level control with GlucoTouch™, respectively a low, normal and high level controls with Super Glucogard II™ (**III**).

4.4 Reference materials

A standard reference material, SRM 909b (National Institute of Standards & Technology, NIST, Gaithersburg, MD 20899, U.S.A.) was used as an internal quality control sample (**I**) and a reference material to evaluate the bias (**III**). SRM 917a *D*-glucose from NIST was applied in spiking of whole blood to attain a hyperglycemic level (**III**). Lithium chloride, p.a. 99% purity (Merck & Co., Inc., Darmstadt, Germany) was used in the preparation of a 50 mM solution, then serially diluted and used for linearity testing in a range between 0.06 mmol/l and 4.01 mmol/l of lithium (**V**).

5. Software applications, statistical methods and calculations

The GUM Workbench® software, version 1.3 (Metrodata GmbH, Grenzach-Wyhlen, Germany) was used to facilitate the calculations of the combined standard uncertainties, u_c and expanded uncertainties U (**II**).

The Analyse-It® with Microsoft Excel 5.0 for Windows software (Analyse-It Software Ltd., 40 Castle Ings Gardens, Leeds, U.K.) was used to:

- test the linearity by ordinary linear regression (**V**)
- test the normality of the laboratory test result data by Shapiro-Wilk W test (**III, V**)
- judge the agreement between the studied method setups with Altman-Bland plots (**V**)
- compare between methods by Passing-Bablok regression analysis (**V**).

The Marchandise equation was applied in the evaluation of the biases from the PT outcome (**V**).

The significant changes expressed in mmol/l were calculated using the generally applicable quality specifications based on biological variation and subject-based reference intervals (**III**).

SPSS® for Windows version 8.0 (SPSS Inc., Chicago, IL, U.S.A) was used to:

- test the intra- and inter-observer variations by the Linear-by-linear Association (**IV**)
- assess the effect of intra-uterine device, patient age and hormone replacement treatment on re-screening parameters (**IV**)
- evaluate the re-viewing agreement between pathologists by cross-tabulating (**IV**).

Microsoft® Excel for Windows version 5.0 (Microsoft Corporation, CA, U.S.A.) was for all basic calculations and tabulating (**I-VI**).

6. Co-operating laboratories, institutes and commercial companies

Medical laboratories in 21 Finnish hospitals and clinical institutes participated voluntarily in the QSL-Finland study (**I**). Other essential co-operators were:

- Helsinki University Central Hospital, Department of Clinical Chemistry (Helsinki, Finland) (**I**)
- Laboratories for Analytical Chemistry and Medical Biochemistry and Clinical Analysis, Faculty of Pharmaceutical Sciences, University of Gent (Gent, Belgium) (**I**)
- Diacor, The Deaconess Institute Clinical Laboratory (Helsinki, Finland) (**I, III**)
- Medix Laboratories Ltd. (Espoo, Finland) (**II, IV, V, VI**)
- Kanta-Häme Central Hospital Laboratory (Hämeenlinna, Finland) (**III**)
- Rinnekoti Foundation Laboratory (Espoo, Finland) (**V**)
- IRMM, Joint Research Centre, European Commission (Geel, Belgium) (**II**)
- Roche Diagnostics (Mannheim, Germany and Espoo, Finland) (**II**).

7. International standards and guides

Outlines of pre-analytic and analytic technical and managerial procedures were tangent to general requirements set to competent testing laboratories in the ISO Guide 25[67] and the standard EN 45001[70] (**I, III, IV, V, VI**). The principles laid down in the EURACHEM/CITAC Guide[115] were followed in the evaluation of the uncertainty of measurement (**II**). Good laboratory practice was followed in all experimental work (**I-VI**).

RESULTS

1. Primary results from the use of the tools I-VI

1.1 Use of reference methods and single donation sera (Tool **I**)

Total calcium and glucose

The blood donors had a healthy background and they were fasting prior to blood donation, which resulted in normal and quite similar concentration levels of total calcium and glucose. The reference method values with uncertainties (Confidence Interval, C.I. 95%) for the six single donation sera are summarized below in Table 1. Ion chromatography was operated under the condition of a maximum bias of 0.7% and ID-GC-MS under the condition of a maximum bias of 0.9%.

Table 1. Target values for total calcium and glucose.

	Total calcium concentration and uncertainty per sample	**Glucose concentration and uncertainty per sample**
Serum sample ID	Mean (mmol/l) (C.I. 95%)	Mean (mmol/l) (C.I. 95%)
ST66	2.340 (2.319 to 2.361)	4.706 (4.682 to 4.729)
H54	2.358 (2.332 to 2.384)	5.107 (4.995 to 5.219)
EP49	2.371 (2.350 to 2.392)	5.779 (5.658 to 5.900)
SR57	2.422 (2.390 to 2.453)	5.719 (5.616 to 5.822)
ST73	2.486 (2.464 to 2.508)	5.995 (5.959 to 6.031)
HX97	2.561 (2.533 to 2.589)	6.279 (6.216 to 6.342)

The imprecision of the total calcium and glucose measurements performed by the QSL-Finland -study participating laboratories were calculated from the mean values derived from two duplicate measurements on three consecutive analyzing days. The number of reported total calcium and glucose measurements was 27 each. Within the laboratories,

the mean CV% varied from 0.2% to 4.4% in total calcium measurements, respectively from 0.2% to 5.2% in glucose measurements.

The deviations from the target values were calculated as mean bias percentages per laboratory. The lowest mean bias% found in total calcium measurements was -0.7% while the highest mean bias% was 5.0%. As glucose was measured, the lowest mean deviation from the reference method value was 0.0% and the highest 3.7%.

1.2 Evaluation of the uncertainty of measurement (Tool **II**)

The results from the use of this tool are based on the quantities and their uncertainties listed in Table 2.

Table 2. List of quantities with units and descriptions.

Quantity	Unit	Description
c'_x	mmol/l	The measurand, i.e. substance concentration of calcium (c_{Ca}) or glucose (c_{Gluc}) in the serum sample
c_x	mmol/l	Substance concentration of calcium or glucose in the sample solution
c_0	mmol/l	Substance concentration of calcium (or glucose) in solution used to establish the zero-point of the calibration curve
c_{cal}	mmol/l	Substance concentration of calcium or glucose in calibrator
A_s	AU[i]	Normalized[ii] and blank-corrected absorbance signal of sample solution in the cuvette
A_0	AU	Absorbance signal from solution used to establish the zero-point of the calibration curve
A_{cal}	AU	Normalized and blank-corrected absorbance signal of calibrator solution in the cuvette
d		Factor describing the possible sample dilution either by the analyst or the instrument
k_{matrix}		Factor describing the contribution from possible matrix effects (difference in commutability of sample and calibrator)
k_{drift}		Factor describing the contribution from an allowed drift in instrument sensitivity
k_{pre}	mmol/l	Term describing the contribution from pre-analytical work (losses, contamination etc.)
k_{intra}		Factor describing the contribution from the intra-individual biological variation of substance concentration of calcium (or glucose) in serum
V_2	µl	The volume of the diluent
V_1	µl	The volume of the sample to be diluted

[i] AU is absorbance units
[ii] Normalized absorbance signal is corrected for absorbances of an empty cuvette and a water-filled cuvette at the requested wavelength(s).

Total calcium

Three cases were studied:

- Case 1 - Uncertainty sources of the analytical phase were considered.
- Case 2 - Also the pre-analytical work (sampling, sample pretreatment and storage) was considered.
- Case 3 - The patient-related uncertainty source (intra-individual biological variation) was added in the evaluation.

The relative contributions of the uncertainty components per cases were expressed as index percentages. The higher the index was, the stronger was the contribution to the evaluated uncertainty budget. Both in case 1 and case 2, the standard uncertainties of the normalized and blank-corrected absorbance signals of sample solution and calibrator in cuvette and substance concentration of calcium in calibrator solution had the strongest contributions to the combined standard uncertainties. In case 3, the intra-individual effect had a dominant influence to the uncertainty budget (Figure 2).

The results from the evaluation of the uncertainty of total calcium measurement in serum were calculated as combined standard uncertainties and relative expanded uncertainties both expressed in mmol/l and percentages (Table 3).

Table 3. Results (mmol/l), combined standard uncertainties (mmol/l and %) and relative expanded uncertainties (mmol/l and %) for the substance concentration of total calcium.

Calcium			
Results			
	Case 1	**Case 2**	**Case 3**
Substance concentration of Ca, c'_x (mmol/l)	2.530	2.530	2.530
Combined standard uncertainty, $u_c(c'_x)$ (mmol/l and %)	0.048 (1.9%)	0.051 (2.0%)	0.070 (2.8%)
Relative expanded uncertainty (k=2), $U(c'_x)$ (mmol/l and %)	0.096 (3.8%)	0.102 (4.0%)	0.140 (5.6%)

Glucose

Three cases were studied:

- Case 1 - Uncertainty sources of the analytical phase were considered. A sample dilution outside the measurement device was considered in the uncertainty evaluation.
- Case 2 - Also the pre-analytical work (sampling, sample pretreatment and storage) was considered. A sample dilution outside measurement device was added in the evaluation.

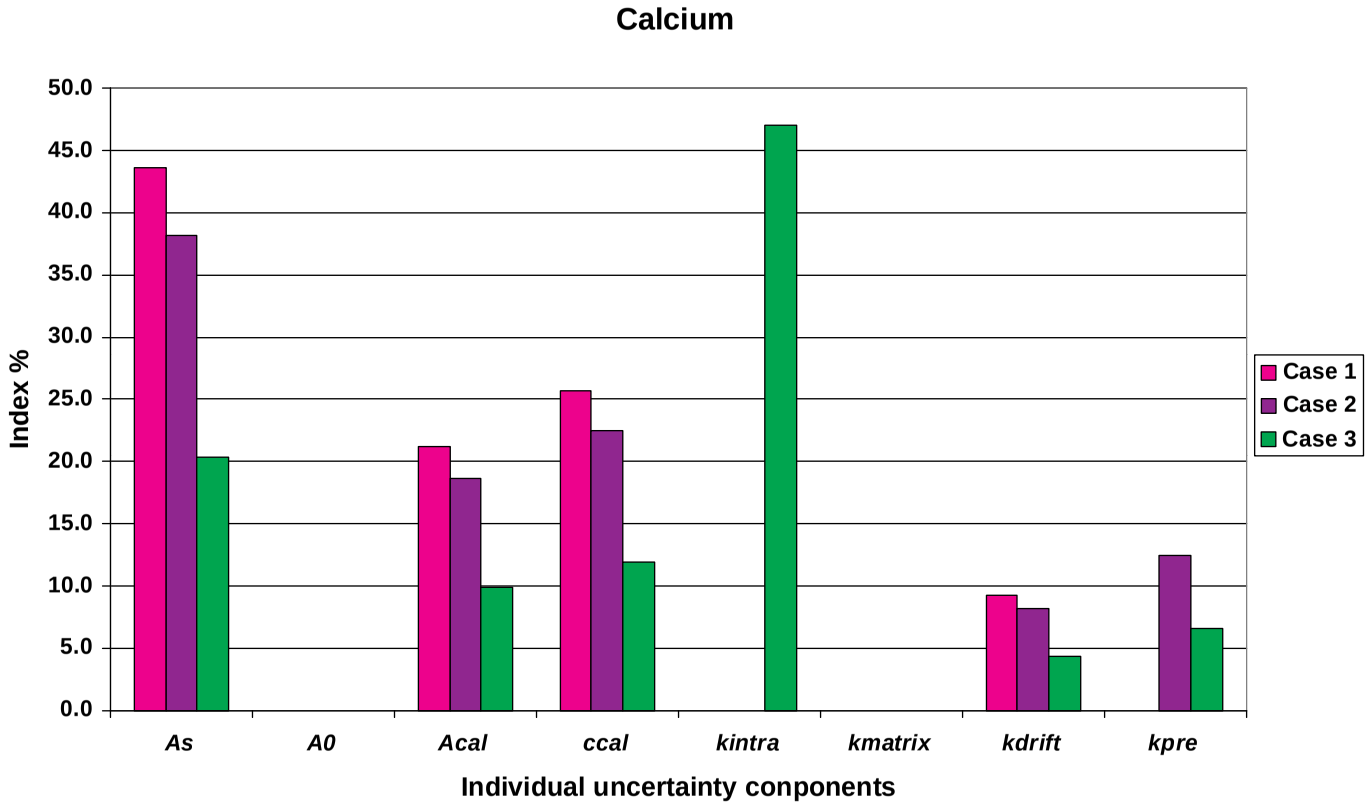


Figure 2. Relative contributions of the individual uncertainty components to the combined standard uncertainty in the measurement of total calcium in serum in the three studied cases

- Case 3 - The patient-related uncertainty source (intra-individual biological variation) was added in the evaluation. The measured glucose concentration fell within the expected range and no dilution was included in the analytical phase.

In case 1 and case 2, the uncertainty sources from the sample dilution and the allowed drift in the instrument sensitivity had the strongest influence to the uncertainty budgets. A distinct finding from the strong effect of intra-individual biological variation related to other uncertainty sources was made (Figure 3). The uncertainty budgets of the evaluation of the uncertainty of glucose measurement in serum in the three studied cases are summarized in Table 4.

Table 4. Results (mmol/l), combined standard uncertainties (mmol/l and %) and relative expanded uncertainties (mmol/l and %) for the substance concentration of glucose.

Glucose			
Results			
	Case 1	**Case 2**	**Case 3**
Substance concentration of glucose, c'_x (mmol/l)	45.83	45.83	6.027
Combined standard uncertainty, $u_c(c'_x)$ (mmol/l and %)	0.562 (1.2%)	0.580 (1.3%)	0.421 (7.0%)
Relative expanded uncertainty (k=2), $U(c'_x)$ (mmol/l and %)	1.12 (2.4%)	1.16 (2.6%)	0.842 (5.6%)

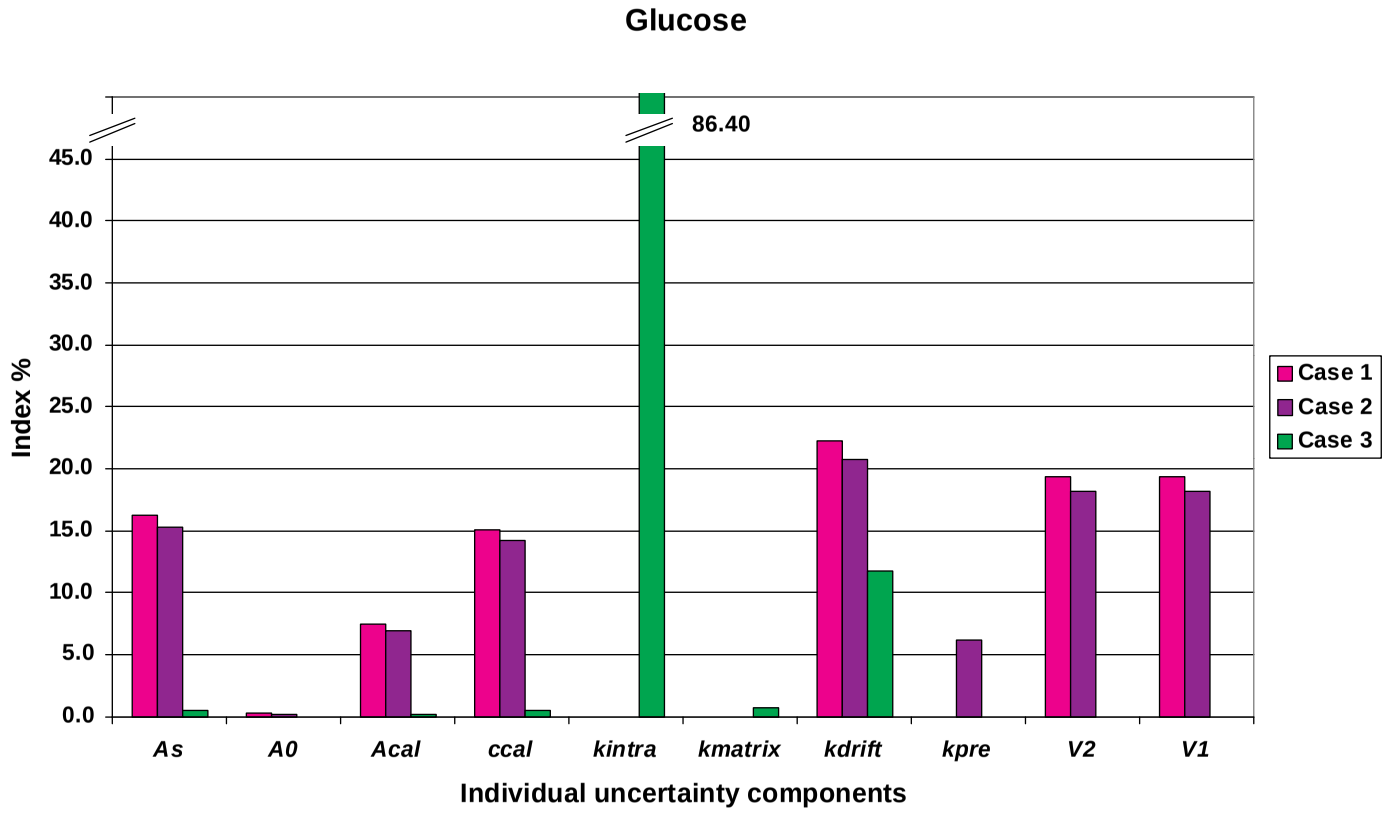


Figure 3. Relative contributions of the individual uncertainty components to the combined standard uncertainty in the measurement of glucose in serum in the three studied cases

1.3 Assessment of performance characteristics (Tool **III**)

The analytical performance and suitability for the intended use of two POCT glucose meters, GlucoTouch™ and Super Glucogard II™, were studied. The outcome was compared to the specifications informed by the manufacturers and/or to medical needs (Table 5).

1.4 Internal quality control (Tool **IV**)

Primary double screening by cytotechnologists

The number of double-screened Pap smears out of the total number of investigated Pap smears was 119/87409 during 1996-1999 at Medix. In the gynecological cytology IQC processes following parameters were screened and evaluated:

- adequacy of diagnostic cellular material
- microbiological findings
- hormonal effects
- leukocyte count
- inflammatory reaction
- cellular atypia
- Papanicolaou classification.

First, excellent inter-observer (n=5) correlation was found in the primary double screening. This resulted from the following findings during 1996-1999:

- estimation of the adequacy of diagnostic cellular material - 99% inter-observer agreement
- estimation of microbiological flora - 95% inter-observer agreement

Table 5. Comparison of performance characteristics of the two POCT meters according to the evaluation.

Performance characteristics	**Glucose level**	**Outcome from the evaluation**			
		GlucoTouch™		**Super Glucogard II™**	
		Manufacturer´s specifications	**Fulfillment of assigned specifications or medical needs**[i]	**Manufacturer´s specifications**	**Fulfillment of assigned specifications or medical needs**[i]
Precision	Hypoglycemic I	Assigned	No	Not assigned	Not tested[ii]
	Hypoglycemic II	Assigned	No	Not assigned	Not compared[iii]
	Euglycemic	Assigned	Yes	Assigned	Yes
	Hyperglycemic	Assigned	Yes	Assigned	No
Variation between test strip lots	Hypoglycemic I	Not assigned	Not acceptable	Not assigned	Not evaluated[iv]
	Hypoglycemic II		Acceptable +/-		Not acceptable
	Euglycemic		Acceptable		Acceptable +/-
	Hyperglycemic		Acceptable		Acceptable
Bias% from plasma level (CLVs[v])	All levels	Not assigned	Acceptable or elevated	Not assigned	Not acceptable
Suitability to neonate care units	Hypoglycemic I and II	Assigned	Suitable	Assigned	Not suitable

[i] Analytical quality specifications based on biology.
[ii] The hypoglycemic level I was beyond the measurement range of Super Glucogard II™ .
[iii] The precision of Super Glucogard II™ at the hypoglycemic level II could not be compared due to the lack of manufacturer´s specifications.
[iv] The variation between Super Glucogard II™ test strip lots at the hypoglycemic level I could not be evaluated due to the limited measurement range
[v] CLVs equals to Clinical Laboratory Values.

- estimation of hormonal effect - positive trend in inter-observer agreement, although non-significant ($p = 0.050$)
- less than 10% intra-observer major disagreements
- significant improvement of accuracy in the inter-observer estimation of inflammatory findings ($p = 0.001$).

Secondly, this study revealed significant individual differences in evaluating cellular atypia, failure and disagreement in benign atypia estimations and failure in primary classification as Papanicolaou class 2 versus class 3.

The affects of intra-uterine device, patient age and hormone replacement treatment on the cytology outcome in Pap smears were categorized as (1) major disagreement, (2) minor disagreement and (3) full agreement (Table 6).

Table 6. Summary of the effects of the intra-individual factors affecting on cytology outcome in primary double screening of Pap smears.

Intra-uterine device	**Effect on cytology outcome**
Leukocyte count	No ($p > 0.050$)
Inflammatory reaction	No ($p > 0.050$)
Cellular atypia	Yes ($p = 0.001$)
Age	
Hormonal effects	Yes ($p = 0.013$)
Leukocyte count	No ($p > 0.050$)
Inflammatory reaction	No ($p > 0.050$)
Cellular atypia	No ($p > 0.050$)
Hormone replacement treatment	
Hormonal effects	Yes ($p = 0.013$)
Inflammatory reaction	Yes ($p = 0.044$)
Cellular atypia	Yes ($p = 0.006$)

Review of the pathologists

The number of reviewed Pap smears out of the total number of investigated Pap smears was 354/87409 during 1996-1999 at Medix. Three of eight senior pathologists reviewed 75% of the quality control Pap smears. This internal quality control process covered re-evaluation of the subsequent parameters with following results between the senior pathologists:

- estimation of good quality in staining – 99% full agreement
- evaluation of the quality of primary screening – 80% full agreement
- assessment of findings and conclusions – 77% full agreement
- assessment of Papanicolaou class – 99% full agreement
- estimation of delay in reporting – 71% full agreement.

1.5 Method validation and result level verification (Tool **V**)

The linearity test of the new ion selective electrode method was accepted at a concentration range from 0.10 mmol/l to 4.00 mmol/l of lithium in serum. The measurement range for lithium reported by the manufacturer was thus verified. During the method validation, the inter-assay variation of the proposed method was superior to the specifications given by the manufacturer. At the six-month checkpoint, higher imprecision was found. The relative bias percentages from EQA consensus mean values, were matched as investigated during the method validation and during the following six months after acceptance to routine analyses. As two EQA samples were analyzed, the results with the new ion selective electrode method deviated more from two reference method values established by flame emission photometry than from the corresponding consensus mean values (Table 7).

Table 7. Results from the tested validation parameters of the three ion-selective electrode setups

Validation parameter	**CobasÒ Integra 700 ISE module / Lab A**	**Microlyte 6 ISE analyzer / Lab B**	**Chiron 654 $Na^+/K^+/Li^+$ analyzer / Lab C**
Linearity	y = 0.970x + 0.014, R^2=0.9996; intercept = 0.014 (95% C.I.: -0.035 to 0.062; slope = 0.970 (95% C.I.: 0.949 to 0.990)	Not tested[i]	Not tested[i]
Measurement range reported by the manufacturer (mmol/l)	0.10 - 4.00	0.20 - 4.00	0.20 - 5.00
Inter-assay variation (CV%) - *method validation*	Low level: 1.7 Therapeutic level: 2.4 High level: 1.5	Therapeutic level: reported as less than 2.0	Low level: 0.91
Inter-assay variation (CV%) - *six month check-point*	Low level: 3.3 Therapeutic level: 2.8 High level: 4.5	Not evaluated[ii]	No evaluated[ii]
Relative bias% from EQA results (CMV[iv])	3.3 (n=6)	5.7 (n=6)	7.3 (n=6)
Relative bias% from EQA results (RMVs[v])	Survey A RMV: 8.7 CMV: 6.1 Survey B RMV: 3.3 CMV: 2.5	Not tested[iii]	Not tested[iii]

[i] The linearity tests for Microlyte 6 ISE and Chiron 654 $Na^+/K^+/Li^+$ analyzers were not specified as validation parameters according to the planned protocol.
[ii] End of subcontracting (Lab B) and withdrawal from the routine analyses (Lab C).
[ii] No EQA results available
[iv] CMV equals to Consensus Mean Value
[v] RMVs equals to Reference Method Values.

1.6 Internal audits (Tool **VI**)

The interviewed medical laboratory personnel had on average a long history with the same employer. Seventy three per cent had worked at Medix for more than 10 years. One of the main findings from the questionnaire revealed that the majority of those who had no previous experience of performing internal audits (83%) were not willing to participate in these quality actions. More than half of the interviewed (66%) stated that the training obtained for internal audits was adequate and that the time put to audits was sufficient (69%). There were strong opinions (86%) about the busy working environment during the audit events. Audit programs were considered to be quite suitable (69%), but even 18% of the interviewed laboratory workers wished for more audit events. Twenty three percent of the workers were unpleased with the information obtained about the outcome of the audits.

2. Results across the set of tools

Three common features with consistent effectiveness were made in the use of the six tools (Table 8):

- demonstration and importance of laboratory performance
- demonstration and assessment of performance characteristics
- importance of personnel skills and supervision.

Firstly, laboratory performance was categorized: (1) demonstrated, (2) indirectly demonstrated, or (3) not demonstrated. Secondly, performance characteristics were either (1) demonstrated (shown) or (2) not demonstrated. Thirdly, personnel skills and supervision was ranked as either (1) important or (2) needed (Table 8).

Essential need for the use of international standards and guides in addition to the compliance with directives was detected.

Traceability was directly associated to the use of reference methods and single donation sera (Tool **I**) and to the evaluation of the uncertainty of measurement (Tool **II)**. Non-traceability and inaccuracy were very evident when the performance of the two POCT glucose meters was estimated (Tool **III**). Complying with quality assurance processes in the frame of international standards and guides became essential due to lack of traceability in pattern recognition and subjective interpretation (Tool **IV**). The novel ion selective electrode method was validated and the result level two-directionally verified with well-established method validation procedure, although traceability could not be indicated (Tool **V**). Managerial laboratory performance could be improved according the outcome of the questionnaire on internal audits, even if the audits, as a part of quality actions, had been regularly and efficiently performed (Tool **VI**).

Table 8. Common features and their effectiveness on the use of the set of Tools **I-VI.**

Set of tools	**Laboratory performance**	**Performance characteristics**	**Personnel skills and supervision**
Use of reference methods and single donation sera (Tool **I**)	Demonstrated	Demonstrated	Important
Evaluation of the uncertainty of measurement (Tool **II**)	Indirectly demonstrated	Demonstrated	Needed
Assessment of performance characteristics (Tool **III**)	Not demonstrated	Demonstrated	Important
Internal quality control (Tool **IV**)	Demonstrated	Demonstrated	Important
Method validation and result level verification (Tool **V**)	Indirectly demonstrated	Demonstrated	Needed
Internal audits (Tool **VI**)	Indirectly demonstrated	Not demonstrated	Important

DISCUSSION

An accuracy-based measurement system stresses traceable measurements when possible. Today, the variety of analytes and measurements systems in the special fields of laboratory medicine cannot guarantee metrological traceability whether it would be possible or not. This makes the availability of international standards and guides important. Table 9. summarizes the outcome of the use of the set of tools.

Table 9. Outcome from the use of the set of tools **I-VI**.

Set of tools	**Traceability**	**International standards and guides**
Use of reference methods and single donation sera (Tool **I**)	Demonstrated	Needed
Evaluation of the uncertainty of measurement (Tool **II**)	Demonstrated	Needed
Assessment of performance characteristics (Tool **III**)	Not demonstrated	Needed
Internal quality control (Tool **IV**)	Not demonstrated	Needed
Method validation and result level verification (Tool **V**)	Not demonstrated	Needed
Internal audits (Tool **VI**)	Not demonstrated	Needed

I Tool: Use of reference methods and single donation sera

The use of Tool **I** concerns the assessment of the-state-of-the-art accuracy of serum total calcium and glucose measurements in a shot of Finnish medical laboratories. The two analytes were chosen to exemplify medical importance, availability of reference methods, and standardized routine methods. Single serum donations served as testing material (Table 8.).

The set-up of the QSL-Finland -study included two primary utensils. The first one being the utilization of reference methods with high accuracy and certified reference materials in producing reference method values. This was to ensure the unbroken traceability chain between field methods and definitive/reference methods as described for an accuracy-based measurement system[9, 25] (Table 9.). Ion chromatography, as potential reference methodology was used for the determination of total calcium[35, 36], respectively ID-GC-MS for the determination of glucose[33, 34].

Secondly, the arrangements of the QSL-Finland -study based on sera from single donations to minimize the possible matrix effects often met with control samples in EQA schemes[160, 33] (Table 8.). Optimally, inter-laboratory comparisons would utilize commutable reference material with traceable target values in assessing the comparability between laboratories and the difference between the observed values and the true value[23]. Therefore, the biases and comparisons to current quality goals were evaluated on the basis of true values although with limited concentration ranges of both analytes. This may have resulted in under-estimated biases, because the samples represented normal values, where the calibration function is at its best. As such, the current study represented a small-scaled, but compact national inter-laboratory comparison with an optimal set-up, even though not so convenient to reorganize more frequently.

The method and calibrator assortment was broad in the measurements of total calcium and glucose among the QSL-Finland participants. This was not a surprising finding because typically numerous different measurement systems (methods, calibrators and instruments) do exist for each measurand in clinical chemistry. System-independent

calibration was found in both total calcium and glucose methods. This diversity of calibration set-ups may indicate either lack of faith in the accuracy of the methods or purposeful minimization of bias (Table 8).

The measurement quality of both analytes was assessed in the co-operating laboratories. EGE-Lab recommends applying the biological approach for analytical specifications[136] (Table 9.). In 1999, the outcome from the Stockholm Consensus Conference strongly advocated this approach[157]. With the exception of four out of 27 reported results, the state-of the-art quality goal for total calcium, imprecision, 1.5% derived from biology was fulfilled. Large biases in reported calcium results predicted problems among many participants. This might be due to unsuccessful method calibration, method unspecificity, or use of empirical factors. Total allowable error of total calcium measurements, both the most stringent derived from biology, 2.5% and the national EQA limit, 3.0% was exceeded by all laboratories with one exception. The superiority of accredited laboratories could not be praised because one accredited method exceeded even the German limit, 10%.

There seemed to be no major problems with the trueness and precision of glucose measurements, because only seven methods resulted in a range beyond the proposed limits derived from biology: 1.9% for trueness and 2.2% for precision. Repeatedly, a single accredited method once more was above the others in bad performance as precision was assessed.

The use of Tool **I** indicated cross-sectional laboratory performance in Finland (Table 8.). The evaluation of performance characteristics in terms of bias and imprecision

showed a need for updating the routine quality assurance procedures by supervision, such as well-established IQC (Table 9.). In this context, fulfilling the requirements of international standards in the sense of accreditation did not give any indication as to be a guarantee for good measurement quality. Analogous conclusions were made from the IRMM International Measurement Evaluation Programme, IMEP-7: Inorganic components in human serum[133].

II Tool: Evaluation of the uncertainty of measurement

The aim of the use of Tool **II** was to emphasize the need for expressing measurement quality in terms of measurement reliability or uncertainty. Evaluation of uncertainty of measurement of total calcium and glucose determinations was exemplified and laboratory performance indirectly demonstrated (Table 8.). At present, an increasing pressure exists for medical laboratories to express and release information of their measurement quality[109, 119]. Evaluation of uncertainty of measurement offers excellent means for this where possible. The EURACHEM/CITAC Guide four-step-procedure was followed in this study[115] (Table 9.).

In addition to regular and special EQA schemes available for medical laboratories, the role of IRMM has been significant in organizing international inter-laboratory comparisons in the frame of IMEP of which the seventeenth round, IMEP-17 has been presently on-going[132]. The close co-operation between IRMM, the European Committee for External Quality Assessment Programmes in Laboratory Medicine (EQALM) and the Committee for Analytical Quality (C-AQ) of the IFCC resulted in a conclusion that there is a need for education of the uncertainty evaluation among laboratory professionals (Table 8.). As the IMEP programs strongly focus to traceability,

uncertainty, use of metrological reference values, traceable to SI system of units, the use of Tool **II** was further supported (Table 9.).

The evaluation process applies data from observations and the calibration function, which is inherently related to quantitative analyses. This emphasizes the character of the calibration and/or reference material, its traceability, and uncertainty. The Essential requirements, Annex 1 in the EC directive set for IVDs gives notice of the requirements for manufacturers about the traceability of values assigned to calibrators and/or control values[12] (Table 9.).

Until today, extremely limited information about the uncertainty of the calibrators used in the field methods has been available for laboratory professionals. This information about the uncertainty of the total calcium and glucose calibrator, and other investigated uncertainty sources was obtained from the manufacturer of the studied analyzing system (Table 9.). The calibration function is crucial in quantitative analyses. The findings from the uncertainty budgeting of total calcium determination gave support to this. The standard uncertainty of the substance concentration of the calcium calibrator showed a significant relative contribution to the combined standard uncertainty as the individual components were investigated (Figure 2). This was also true with glucose measurement in Cases 1 and 2 (Figure 3).

Medical laboratory analyses consist of several phases and thus include a number of possible uncertainty sources, major or minor. The four-step-procedure was found to be an excellent method of iterating the entire measurement process. Critical examination of each single phase results in the best possible evaluation of uncertainty. It is intended

that data on observations from IQC and method validation are used in the evaluation process and this data is usually available. As it was with this study, data from literature may also be applied.

Interestingly, in the pre-dilution phase the volume of the diluent and the volume of the sample to be diluted had major relative contribution to the combined standard uncertainty (Figure 3).

It may be claimed that pre-analytical factors such, as intra-individual biological variation should be beyond the scope of uncertainty budgeting when measurements are involved. However, adding this factor to the other quantities resulted in a surprisingly large relative contribution in both determinations (Figure 2 and Figure 3). Consequently, it is of utmost importance to set quality goals, and they should be based on biological variation.

The outcome of the use of Tool **II** was an example of the usefulness of the evaluation of uncertainty of measurement in routine clinical chemistry. Evaluated uncertainty of measurement as a measure of reliability is a valuable performance characteristic and it can be indirectly used when measuring laboratory performance, i.e. assessing the success in EQA schemes. The findings from the use of Tool **II** support this statement. Much of the purpose of this study was educational and it is argued that practical guidelines are still needed (Table 9).

III Tool: Assessment of performance characteristics

IVDs for SMBG belong to List B of diagnostic devices according to Annex 2 in the Directive 98/79/EC on IVDs[12]. It stipulates that calibrators and reference materials of SMBG meters belong to the legislative inspection and approval whenever trading and usage for monitoring purposes is concerned in the European Community. In the US, CLIA'88 classifies SMBG meters into the category of waived tests[64, 97]. More than two hundred glucose monitoring devices, including test strips are listed in the database of FDA's Center for Devices and Radiological Health[161]. This supports the essential role of glucose meters in the family of all POCT systems.

In addition to self-monitoring, blood glucose meters are widely used in hospitals, health care centers, and doctors' offices. Quantitative results are produced and used for important decision-making. For this reason, analytical performance and fit-for-purpose evaluation of the device is of the utmost importance. Generally, analytical performance should not be judged if quality goals do not exist. The Directive 98/79/EC defines the essential criteria while clinical chemistry expertise is needed for determination of performance criteria[154, 155].

By investigating any analyzing system, performance characteristics, even selected ones may be evaluated and assessed (Table 8.). The use of Tool **III** was shown with the comparison of two SMBG meters, the GlucoTouch™ and the Super Glucogard II™ (Table 5.). The evaluation was based on both self-made observations and manufacturers' specifications. Laboratory performance *in situ*, was beyond the scope of this study, although uncertainty of measurement was discussed when clinical laboratory

values were concerned (Table 8.). In any case, one might regard testing at the hospital wards as satellite laboratory work.

Reference methods and reference materials have been available for the measurement of glucose for decades, but the traceability of the two SMBG test systems remained undesignated (Table 9.). No such data was available from either of the manufacturers or their representatives. Dedicated glucose analyzing systems for the calibration of SMBG meters exists, but traceability is scarcely discussed[162, 163.] Under the circumstances of this study, the calibration data of both meters was insufficient. The GlucoTouch™ did measure the plasma glucose level as assigned by the manufacturer, while the trueness of Super Glucogard II™'s especially at the hypoglycemic levels remained quite dubious.

ISO quality system standards set the requirements for companies that wish to conform to the standards and certification. The disciplined design control may not have been in place if the intended use and the performance of a POCT device did not fulfill the current requirement[92]. Pitfalls may exist in the family of SMBG meters and the current example of the use of Tool **III** certainty demonstrated it. Fortunately, the Directive 98/79/EC on IVDs has come into force and will filter the poorly performing newcomers out of this pool of very diverse analyzing systems. International standards strengthen the triangle of IVD end-users, manufacturer's and official controlling bodies.

IV **Tool:** Internal quality control

IQC in gynecological cytopathology was included in the set of investigated quality tools to represent a field of laboratory medicine where traceability to SI cannot be shown.

Approximately 95% of all medical laboratory testing still falls in the range of metrological untraceability[39] and much of it will remain in this group as e.g. cellular testing material is often concerned. Laboratory accreditation according to the current standard, ISO/IEC 17025[84] strongly advocates traceability to SI where possible. As cytopathology and Pap smear testing are concerned, other means of demonstrating the measurement quality and laboratory performance or competence need to be used (Table 9.).

The methods of IQC in gynecological cytopathology, double screening and pathologist reviewing differ from those of the traditional ones, mostly related to clinical chemistry and more familiars to clinical biochemists. In fact, error detection and error prevention are the basic goals in internal quality assurance procedures regardless of the specific field of laboratory medicine. Primary screening and investigation of all testing material by a senior pathologist was a standard testing procedure from 1996-1999 at Medix. Regarding error detection, Tool **IV** was exemplified by factors affecting the measurement quality of primary screening.

The laboratory report of a Pap test result contains several parameters first pre-screened and then investigated by pathologist. Seven parameters belonged to the standard testing procedure. Reference methodology, although not a primary method of measurement by definition, the conventional Papanicolaou's staining and cell morphology based on the Bethesda System[42] was used and regarded as the best possible one for routine analyses (Table 9.). Inter-observer correlation and review of the pathologists were investigated as performance characteristics and indications of measurement quality. The inter-observer (n=5) correlation was excellent. A positive trend with time although

statistically non-significant in inter-observer agreement on the estimation of hormonal effect was found a consequence of the feedback meetings between senior pathologists and cytotechnologists. A similar reason for the improvement of the accuracy in the inter-observer estimation of inflammatory findings was concluded.

The IQC test material was grouped in terms of intra-individual characteristics, such as intra-uterine device, age and hormone replacement treatment to find out the effect of these factors on the primary double screening. Major and minor disagreements, and full agreements were evaluated (Table 6.). Less effect than expected was found. Nevertheless, hormone replacement treatment and the presence of intra-uterine device effected significantly on pre-screening of cellular atypia. It was not surprising that the results from the investigation of patient's age and the evaluation of hormonal effects strongly indicated the difficulty of screening of this parameter among patients younger than 47 years.

The pathologists reviewed 354 of 87409 Pap smears between 1996-1999 at Medix. Out of this reviewed sample pool, three of the eight senior pathologists re-evaluated 75% of the IQC samples. The low number of the others' reviews made the statistical calculations impossible. Anyway, they agreed on the quality of the stained Pap smears (99%) and primary screening (80%). This is a sign of good laboratory practices and skillful cytotechnologists. Reviewing by senior pathologists is always done long after the laboratory reporting. In this context, possible error detection does not occur in an optimal time. On the other hand, evaluation of this performance characteristic should result in harmonization of subjective interpretation and statements resulting to error prevention (Table 9).

V Tool: Method validation and result verification

In the course of a continuously developing medical laboratory environment, method validation and verification of novel or replacement test methods of measurements is persistent work among clinical biochemists and clinicians. Frequently, these kinds of tasks are performed under limited time schedules. Under these circumstances, much weight is placed on established validation and acceptance procedures. Although the first method evaluation schemes were introduced already three decades ago[100], these procedures do not necessarily belong to basic quality tools when implementing quality systems in medical laboratories today.

The aim of using Tool **V** was to introduce a practical validation example taken from routine medical laboratory production. There are four main stages: planning, performing, evaluating and verifying are typical elements in a common validation process. First and not least, it is postulated that methods of measurement should be made to satisfy an agreed requirement[27]. Validation parameters should be defined in each case to fulfill practicability, reliability and cost effectiveness of the procedure. For this reason, the following utensils of Tool **V** were selected as the necessary validation parameters: testing of linearity, verification of measurement range, and evaluation of imprecision and bias.

Verifying by IQC and EQA closes up the validation process[75]. Thus, the laboratory performance was indirectly demonstrated at the 6-month checkpoint by EQA (Table 8.). None of the three manufacturers of the ion selective electrode setups could show metrological traceability (Table 9.). During the planning, a decision was made that

certified reference materials would not be purchased for a single method validation due to their high costs (Table 9.). Fortunately, two EQA schemes offered reference method values on serum lithium at the 6-month checkpoint. The results showed that modeled laboratory performance with EQA past samples was better than at the 6-month checkpoint. This does not naturally make the EQA past samples useless in method validation, but necessary for future comparison.

The main indications of the determination of serum lithium are therapeutic drug monitoring and avoidance of intoxication. As the therapeutic range is narrow for serum lithium, 0.60 - 1.2 mmol/l, excellent precision is required. Verification of the imprecision at the 6-month checkpoint revealed similar finding as with the bias. The observed precision of the novel ion selective electrode method during the method validation was better than assigned by the manufacturer, but similar findings could not be made at the 6-month checkpoint even with the system controls. The precision of the system-independent control was superior to that of system controls, but did not quite fulfill the national criteria of 2%.

The method comparisons with patient samples (n = 62) showed biases between all methods and the CIs of the intercepts did not overlap. This finding was nevertheless in agreement with the outcome from the evaluation of relative biases from consensus mean values although the use of the Marchandise equation loses the sign of bias.

VI Tool: Internal audits

Measurement quality is certainly closely related to technical competence. The requirements in the current standards, ISO/IEC 17025[84] and EN/ISO 15189[85] cover the laboratory management as well. The principles of total quality management stress to the Plan-Do-Check-Adjust -approach[46] and internal audits Tool **VI** should be utilized as an essential management tool. In the frame of laboratory accreditation or certification, internal audit in medical laboratories is a rather new concept and it could be predicted that the character of even this management tool would change with time.

The idea of operating with Tool **VI** was to check and then adjust the established practice of internal audits by finding out about the opinions of the laboratory personnel.

Integrated participation in quality actions is essential. The lack of experience among the personnel to perform internal audits appeared to be the reason for stumbling in the yearly audit planning. It was found that continuous encouragement for auditing had not been sufficient (Table 8.). Despite the existing documented practice, i.e. internal audit as a quality assurance process, the necessity for training and supervision and importance of personnel skills even in this practice was observed (Table 9.).

Nearly one fifth (18%) of the interviewed laboratory workers wished for more audit events than what was normally planned and accomplished, while the rest were satisfied with the current procedures. One could consider this as a positive way of quality thinking. Willingness to improve one's work is a favorable sign towards continuous quality improvement. Less surprising was that more than one fifth (22%) was disappointed with the flow of information regarding the outcome of the internal audits.

SUMMARY AND CONCLUSIONS

Medical laboratories have faced tremendous changes in refining the idea of measurement quality during the last fifteen years. Although quality is not a novel invention, the quality boom has indeed produced international standards, guides, and recommendations to be applied in the field of medical laboratory science. This work continues within several expert groups worldwide. Consequently, a quality-oriented way of laboratory working has changed the general attitude towards transparency, systematic, and traceability by definition.

As a selection of the numerous quality tools that are available, the presented set of Tools **I-VI** indicated practicability and usefulness. Measurement quality was shown in terms of laboratory performance, performance characteristics, personnel skills, and supervision.

The use of the presented set of tools strongly advocates the need and necessity of international standards and guides. It is important to understand that all test methods within the many fields of laboratory medicine should be treated under harmonized rules of quality assurance whenever a quality-oriented way of laboratory working is required.

The use of primary methods of measurements ties the results to SI at the top level of an accuracy-based measurement system. Traceable reference method values offer the best possible means for demonstrating the laboratory performance whenever possible. To avoid high costs reference method values combined with commutable control material should be at least regularly available for as many inter-laboratory comparisons and

EQA schemes as possible in the future. Medical laboratory professionals should pay more attention to the calibration and IQC procedures of their field methods simultaneously as manufacturers are fulfilling the requirements set in the IVD directive. It is concluded that laboratory accreditation in its present form does not ensure good measurement quality.

The importance of traceability and evaluation of uncertainty of measurement are strongly focused upon in the IMEP program. As an essential performance characteristic, uncertainty of measurement in quantitative analyses reflects the measurement quality in the best possible way. The four-step uncertainty evaluation process ensures a thorough investigation of a measurement procedure. Notwithstanding, a brush-up of skills and metrological approach among laboratory professionals is still required. The availability of sophisticated software tools does not diminish this need.

IVDs used for SMBG represent a group of POCT instrumentation by which more on-site decisions are made than with any other analyzing systems. Precision is generally regarded as a more important performance characteristic in monitoring than trueness. Although glucose meters are listed in Annex 2 of the IVD directive, no traceability could be shown. Furthermore, measurement capabilities should comply with the manufacturer's specifications and the intended use with possible limitations should be clearly assigned. It was concluded that co-operative supervision is needed whenever the POCT devices are under the responsibility of central laboratory. Purchasing and performance evaluation combined with IQC should be in the hands of laboratory personnel. The current international standards should strengthen the triangle between IVD end-users, manufacturers and authorities.

In the measurement field of pattern recognition and subjective interpretation, metrological traceability cannot be demonstrated today. Other quality tools, such as IQC are then strongly weighted. Double screening or reviewing just in terms of controlling, does not result in the improvement of diagnostic quality as such. It was found that, feedback meetings resulted in harmonized evaluation of cellular material and error prevention. IQC means extra workload in cytopathology due to its manual and time-consuming nature of analysis. There are good grounds for lowering the frequency of double screening if a high probability of error detection can be demonstrated, if the inter-observer correlation is good, and if the technical competence of the pre-screening has been proved. The outcome of this presented quality tool may be used as a measure of diagnostic reliability at its best.

Laboratory tests should be made to fulfill the needed requirements. That is what analytical specifications are needed for. Well-planned, performed, documented, and verified method validation is a prerequisite of reliable routine methods. The awareness of performance characteristics is essential for re-establishing IQC procedures and assessment of EQA outcome. Validation parameters shall be rationally defined case by case including both standardized and non-standardized test methods. Careful validation of accredited methods should not be kept separate from the non-accredited ones. Manufacturer's specifications can be used as additional and supporting data for validation but not as a sole source of data. The use of EQA past samples where possible offer valuable means in predicting the future success in EQA schemes and in verifying the result level between novel and current methods of measurement. The use

of EQA samples with reference method values can be utilized in the evaluation of bias if available.

Good laboratory practice means re-evaluating the established processes. Today, internal audits are a documented management process in every medical laboratory with a quality system. It was concluded that carrying out and completing this quality action needs updating in its management. Brief questionnaires can be used with good results whenever one wishes the current processes to be improved. Laboratory performance in the meaning of management success could be demonstrated. As a summary, additional training and improving the internal information flow were clearly needed.

ERRATUM

I Linko S, Himberg J-J, Thienpont L, Stöckl D, De Leenheer A. Assessment of the state-of-the-art trueness and precision of serum total-calcium and glucose measurements in Finnish laboratories - the *QSL-Finland -study. Scand J Lab Invest* 1998; 58:229-240.

Then mean CV% per laboratory is missing from the table in the original reprint.

TABLE VI. Imprecision of glucose measurements, CV% (n=6)

Sample	**ST66**	**H54**	**EP49**	**SR57**	**ST73**	**HX97**	
RMV (mmol/l)	4.706	5.197	5.779	5.719	5.995	6.279	Mean CV%
Lab	CV%	CV%	CV%	CV%	CV%	CV%	per laboratory
O	0.8	0.0	0.0	0.0	0.0	0.7	0.2
P	1.6	2.1	1.6	0.9	1.8	1.2	1.5
W	0.8	0.9	0.9	1.2	1.5	0.0	0.9
Y	3.4	2.1	2.7	2.2	2.1	2.4	2.5
B	1.8	0.7	0.0	0.8	0.8	0.8	0.8
K	1.0	0.8	0.9	1.2	0.8	0.7	0.9
V	2.2	2.0	1.6	2.0	1.8	2.6	2.0
X	3.4	2.7	2.4	3.6	4.0	3.7	3.3
A	2.0	1.6	1.3	3.2	0.8	1.5	1.7
Da	1.1	0.7	1.3	0.9	0.8	0.7	0.9
E	1.6	1.9	1.5	2.0	1.8	2.1	1.8
M	2.8	1.7	3.6	3.3	3.4	1.4	2.7
N	1.1	1.1	0.9	0.8	1.0	0.7	0.9
Qa	1.5	1.3	1.8	0.0	0.8	0.9	1.1
Ua	5.2	3.7	5.3	5.6	5.1	6.2	5.2
Ub	1.5	2.1	1.3	1.4	0.8	4.3	1.9
H	1.1	0.9	1.9	0.2	1.2	1.3	1.1
J	1.6	2.7	1.4	1.3	1.2	1.7	1.6
Qb	3.2	1.7	1.3	2.8	3.2	1.5	2.3
S	2.6	1.0	1.5	1.6	2.3	1.3	1.7
L	4.1	2.9	2.7	5.0	3.2	3.7	3.6
T	2.1	2.7	3.1	2.7	1.6	3.0	2.5
G	1.1	1.5	1.2	1.3	1.4	1.3	1.3
R	1.6	2.2	1.7	3.0	0.8	1.2	1.8
C	0.8	0.9	2.1	0.6	0.8	0.6	1.0
Db	0.8	1.4	0.8	1.3	1.1	1.3	1.1
F	0.8	1.8	0.9	0.6	0.6	0.8	0.9

Min $CV\%_{all}$ = 0.0 (n=162); Max $CV\%_{all}$ =6.2 (n=162)

REFERENCES

1 BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. International vocabulary of basic and general terms in metrology (VIM). 2nd ed. Geneva: ISO, 1993.

2 Bureau International des Poids et Mesures (BIPM). Comité Consultatif pour la Quantité de Matière, Proceedings of the 7th meeting of CCQM (2001), 95 pp. Committees, Consultative Committees, Publications of the CCQM. http://www.bipm.fr (November 2002)

3 Bureau International des Poids et Mesures (BIPM). The Convention of the Metre. http://www.bipm.fr (November 2002).

4 Bureau International des Poids et Mesures (BIPM). SI, The International System of Units. 7th ed., Sèvres: BIPM, 1998:152 pp. Official publications of the BIPM. http://www.bipm.fr (November 2002).

5 De Bièvre P, Taylor PDP. Traceability to the SI of amount of substance measurements: from ignoring to realizing, a chemist´s view. Metrologia 1997; 34: 67-75.

6 International Federation of Clinical Chemistry, International Union of Pure and Applied Chemistry. Approved recommendations (1978) on quantities and units in clinical chemistry. Prepared by Dybkær R. Pure Appl Chem 1979; 51: 2451-79.

7 De Bièvre P. The key elements of traceability in chemical measurement: agreed or still under debate? Accred Qual Assur 2000; 5: 423-428.

8 Zschunke A. Global comparability of analytical results. Accred Qual Assur 1998; 3: 393-397.

9 Cali PJ. An idea whose time has come. Clin Chem 1973; 2: 291-3.

10 Tietz NW. A model for a comprehensive measurement system in clinical chemistry. Clin Chem 1979; 25: 833-9.

11 EURACHEM/CITAC Guide Traceability in Chemical Measurement. Workshop Draft June 2002. Guides & Documents. http://www.eurachem.ul.pt/ (November 2002).

12 Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on *in vitro* diagnostic medical devices. Official Journal of the European Communities (1998), L 331/1-37.

13 In Vitro Diagnostic Medical Devices - Measurement of quantities in samples of biological origin - Metrological Traceability of values assigned to calibrators and control materials, ISO/CD 17511. Geneva, ISO, February, 2000.

14 In Vitro Diagnostic Medical Devices - Measurement of quantities in samples of biological origin - Metrological Traceability of values for catalytic concentration of

enzymes assigned to calibrators and control materials, ISO/CD 18153. Geneva, ISO, February, 2000.

15 ISO Guide 30:1992 Terms and definitions used in connection with reference materials. 2nd edition. ISO, Geneva.

16 National Committee for Clinical Laboratory Standards: Terminology and Definitions for use in NCCLS Documents: Proposed Standard. 3rd ed. NCCLS Document NRSCL8-P3. Wayne. PA, National Committee for Cilinical Laboratory Standards, 1996.

17 Büttner J. The need for accuracy in laboratory medicine. Eur J Clin Chem Clin Biochem 1995; 33: 981-988.

18 Büttner J. Referece materials and reference methods in laboratory medicine: a challenge to international cooperation. Eur J Clin Chem Clin Biochem 1994; 8: 571-577.

19 Pauwels J, Lamberty A. CRMs for the 21st century: new demands and challences. Fresenius J Anal Chem 2001; 370: 111-114.

20 Schimmel H, Kristiansen N, Grasserbauer M. Metrological support of the Institute for Reference Materials and Measurements. Clin Chem Lab Med 1999; 37: 485-486.

21 Institute for Reference Materials and Measurements, European Commission, Joint Research Centre, Geel, Belgium. IRMM Reference Materials. http://www.irmm.jrc.be (November 2002).

22 National Institute of Standards and Technology, Gaithersburg MD, U.S.A. Standard Reference Materials. http://www.nist.gov (May 2002).

23 Franzini C. Commutability of reference materials in clinical chemistry. JIFCC 1993; 4: 186-192.

24 Ferard G, Lessinger J-M. Preparation of enzyme calibration materials. Clin Chem Acta 1998; 278: 151-162.

25 Uriano, GA & Cali JP. (1977) Role of reference materials and reference methods in the measurement process. In: Validation of the measurement process (DeVoe, JR, ed.) ACS Symp. Series 63. American Chemical Society, Washington, pp. 140-160.

26 Milton MJ, Quinn TJ. Primary methods for the measurement of amount of substance. Metrologia 2001; 4:289-296.

27 Tietz Textbook of Clinical Chemistry, edited by Burtis CA, Ashwood ER. Chapter 17, Quality Management, Reference materials and calibration, p. 391,321 (1999), 3rd edition. WB Saunders Company. ISBN 0-7216-5610-2.

28 Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998; 6: 449-457.

29 Cali JP, Mandel J, Moore L. (1972) Standard reference materials: a referee method for the determination of calcium in serum, NBS Special Publication 260-36.

30 Velapoli RA, Paule RC, Schaffer R, Mandel J, Machlan LA, Garner EL, Rains TC. (1980) Standard Reference Materials: A reference method for the determination of lithium in serum, NBS Special Publication, 260-69.

31 Hannestad U, Lundblad A. Accurate and precise isotope dilution mass spectrometry method for determining glucose in whole blood. Clin Chem 1997; 5: 794-800.

32 De Bièvre P. Isotope dilution mass spectrometry (IDMS). (1994) Trace element analysis in biological specimens. Techniques and instrumentation in analytical chemistry, vol 15, ed. By Herber RFM and Stoeppler M. Elsevier Publications, pp. 169-183.

33 Stöckl 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; 6: 993-1000.

34 Thienpont L, De Leenheer AP, Stöckl 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; 6: 1001-1006.

35 Thienpont LM, Van Nuwenborg JE, Reinauer H, Stöckl D. Validation of candidate reference methods based on ion chromatography for determination of total sodium, potassium, calcium and magnesium in serum through comparison with flame atomic emission and absorption spectrometry. Clin Biochem 1996; 6: 501-506.

36 Thienpont LM, Van Nuwenborg JE, Stöckl D. Ion chromatography as potential reference methodology for the determination of total calcium and magnesium in human serum. Anal Chem 1994; 14: 2404-2408.

37 EN 12286 (1998) In vitro diagnostic medical devices - Measurement on quantities in samples of biological origin - Presentation of reference measurement procedures. CEN/CENELEC.

38 Louie RF, Tang Z, Shelby DG, Kost GJ. Point-of-care testing: Millenium technology for critical care. Lab Med 2000; 7: 402-408.

39 Neil Greenberg N. Calibrator Traceability: The Industry Impact of the IVD Directive's New Requirements. IVD Technology 2001; 2: 18-27.

40 Bramley R, Brown A, Ellison S, Hardcastle W, Martin A. Qualitative analysis: a guide to best practice-forensic science extension. Sci Justice 2000; 40: 163-70.

41 Ellison SLR. Uncertainties in qualitative testing and analysis. Accred Qual Assur 2000; 5: 346-348.

42 Solomon D, Davey D, Kurman R et al.; The Forum Group Members.; The Bethesda 2001 Workshop. JAMA 2002; 287(16): 2140-41.

43 ISO 8402:1994. Quality management and quality assurance. Vocabulary. ISO, Geneva.

44 BS 4778: Part 2:1991. "Quality vocabulary, part 2 - Quality concepts and related definitions". BSI, London, U.K.

45 Burnett D. Understanding accreditation in laboratory medicine. Chapter 1 What is accreditation? p. 5. Chapter 2 Standards, guidelines, criteria or principles? p. 15,5,26. ABC Venture Publications 1996. ISBN 0902429205.

46 Deming WE. Out of the Crisis. Cambridge: Center for Advanced Engineering Study, Massachusetts Institute of Technology, Cambridge, MA, 1986. ISBN 0-911379-01-0.

47 Lansky DJ. Patients as consumers: Making the health care system our own. Ninth annual Herbert Lourie Memorial Lecture on health policy. No. 13/1998. Center for Policy research, Maxwell School, 426 Eggers Hall, Syracuse, NY 13244-1020.

48 Dorsey DB. Evolving conceps of quality in laboratory practice. Arch Pathol Lab Med 1989; 113: 1329-1334.

49 Levey S, Jennings ER. The use of control charts in the medical laboratory. Am J Clin Pathol 1950; 20: 1059-66.

50 Shewhart WA. Economic control of quality of the manufactured product. Van Nostrand, New York, NY, 1931.

51 Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in medical chemistry. Clin Chem 1981; 27: 483-501.

52 Bremond J, Plebani M. IVD industry role for quality and accreditation in medical laboratories. Clin Chim Acta 2001; 309: 167-171.

53 Kallner A. Quality management in the medical laboratory: a comparison of draft standards. Clin Chem Acta 1998; 278: 111-118.

54 IFCC/EMD/C-AQ: Hill P, Uldall A, Wilding P. Fundamentals for External Quality Assessment. IFCC, London 1996: 1-45.

55 Libeer JCPM. External Quality assessment in Clinical Laboratories. Dissertation, Antwerpen, 1993. ISBN 90-900-6660-0.

56 Westgard JO, Barry PL. Cost-effective quality control: managing the quality and productivity of analytical processes. AACC Press, 1725 K Street NW, Washington DC-20006, 1986.

57 El Nageh MM, Heuck C, Kallner A, Mayard J. Quality systems for medical laboratories. The World Health Organization (1995). WHO Regional Publications, Eastern Mediterranean Series. ISBN 92-9021-203-9.

58 ISO/IEC Guide 2:1991 General terms and their definitions concerning standardization and related activity. ISO, Geneva.

59 The Collins English Dictionary. 3rd Edition, 1994. ISBN 0 00 470678-1.

60 EN 45020:1993 Glossary of terms for standardization and related activities. CEN/CENELEC, Brussels.

61 Locke JW. Development of laboratory accreditation in the United States. Review. Accred Qual Assur 1998; 3: 356-361.

62 Batjer JD. The College of Americal Pathologists laboratory accreditation programme. Clin Lab Haematol 1990; 12 (Suppl 1):135-138.

63 Roberts JS, Coale JG, Redman R. A history of the Joint Commision on Accreditation of Hospitals. JAMA 1987; 258: 936-940.

64 US Department of Health and Human Services (1992) Medicare, Medicaid and CLIA programs: Regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. Fed Regist 57: 7002-186.

65 Davies P. NATA: 50 years experience with accreditation. Accred Qual Assur 1996; 1: 240-246.

66 ISO 9000 / ISO 14000. International Organization for Standardization. Geneva. http://www.iso.ch (November 2002).

67 ISO/IEC Guide 25:1990 General requirements for the competence of calibration and testing laboratories. ISO, Geneva.

68 Canadian Council on Health Service Accreditation: A Client Centered Accreditation Program, 1995 Standards Framework.

69 OECD. Committee for the final report of the working group on the mutual recognition of compliance with good laboratory practice. ENV/CHEM/CM 87.7. OECD Environment Monographs 1988 (no.18): 25-35 & 41-54, Paris.

70 EN 45001:1989 General criteria for the operation of testing laboratories. CEN/CENELEC, Brussels.

71 Kallner A. Accreditation of medical laboratories. Some reflections from the Nordic horizon. Clin Chim Acta 2001; 309: 163-165.

72 Finnish Accreditation Service. Testing laboratories. Clinical Analytics. http://www.finas.fi. (November 2002).

73 EA-European Accreditation. Multilateral Agreement, MLA. http://www.european-accreditation.org. (November 2002).

74 Accreditation Handbook. Version 6.2 Sheffield: Clinical Pathology Accreditation (UK). The Children Hospital, Western Bank S10/2TH; 1995.

75 Dybkær R, Jordal R, Jørgensen PJ, Hansson P, Hjelm M, Kaihola H-L, Kallner A, Rustad P, Uldall A, de Verdier C-H. A quality manual for the clinical laboratory including the elements of a quality system. Proposed guidelines. Scand J Clin Lab Invest 1993; 53 suppl. 212.

76 Suositus kliinisten laboratorioiden laatukäsikirjan laatimiseksi. Version 2.0. Editors S. Linko and L. Linko. Moodi 1999; Special issue No. 8: pp. 25. Labquality, Helsinki, Finland. ISBN 951-96219-8-9.

77 Loeber JG, Slagter S. Code of practice for implementation of a quality system in laboratories in the health care sector (1991), CCKL Secretariat, Bilthoven, Netherlands.

78 Empfelungen für die gute praxis im medizinisch-analytischen laboratorium (GPMAL) 1993, Switzerland.

79 SFBC. Guide de bonne execution des analyses de biologie medicinale (G.B.E.A). Project 1994; 1994, p. 1-28.

80 European cooperation for Accreditation of Laboratories (EAL) (1997) Accreditation for Medical Laboratories. EAL-G25/ECLM-1. Edition 1.

81 WHO/EURO – ECCLS Guidelines on Good Practice in Clinical Laboratories. Part I Clinical chemistry (1991) World Health Organization, Regional Office for Europe. European Committee for Clinical Laboratory Standards.

82 European Communities Confederation of Clinical Chemistry (ECCCC) (1996) Essential Criteria for Quality Systems of Clinical Biochemistry Laboratories in the EU. Working Group on Harmonisation of Quality Systems and Accreditation, EC4.

83 Kociba P, Siloaho M. Kliinisen kemian laboratorioiden akkreditointi Suomessa. Kliin Lab 1997; vsk14:10-12.

84 ISO/IEC 17025:1999 General requirements for the competence of testing and calibration laboratories. ISO, Geneva.

85 EN/ISO 15189. Quality management in the medical laboratory. ISO, Geneva, (2002).

86 Burnett D, Blair C. Standards for the medical laboratory – harmonization and subsidiarity. Clin Chim Acta 2001; 309: 137-145.

87 Kenny D. ISO and CEN documents on quality in medical laboratories. Clin Chim Acta 2001; 309: 121-125.

88 Proposal for another step on the path to total quality management. Kallner A, Waldenstrom J. http://www.ifcc.org/ejifcc/vol13no1/1301200105.htm (May 2002).

89 ISO 9001:1999 Quality systems – model for quality assurance in design/development, production, installation and servicing. ISO, Geneva.

90 ISO 13485:1998 Quality systems – medical devices – particular requirements for the application of ISO 9001. ISO, Geneva.

91 Lasky FD, Boser RB. Designing in quality through design control: a manufacturer's perspective. Clin Chem 1997; 43: 866-872.

92 Lasky FD. Quality planning in the 'in vitro diagnostic` industry. Accred Qual Assur 2001; 6: 415-420.

93 Powers DM, Greenberg N. Development and use of analytical quality specifications in the in vitro diagnostics medical device industry. Scand J Clin Lab Invest 1999; 59: 539-544.

94 Libeer JC, Goldschmidt HMJ. Introduction for a special issue on the Antwerp Conferences. Accr Qual Assur 1999; 4: 76-8.

95 US Food and Drug Administration. Design control guidance for medical device manufacturers. FDA March 11, 1997. http://www.fda.gov/cdrh/guidance.html (May 2002).

96 Global Harmonization Task Force. Design Control guidance for medical device manufacturers. Sept.11, 1997. http://www.ghtf.org (May 2002).

97 Regulations for implementing the Clinical Laboratory Improvement Amendments of 1988: A summary MMWR 41(RR-2); 001. Publication date: 02/28/1992.

98 European cooperation for Accreditation of Laboratories. Validation of test methods. EAL-P11, ed. 1, February 1997.

99 Eurachem Guide The Fitness for Purpose of Analytical Methods. English ed. 1.0, 1998, ISBN: 0-948926-12-0.

100 Barnett RN, Youden WJ. A revised scheme for the comparison of quantitative methods. Am J Clin Pathol 1970; 54: 454-462.

101 Evaluation methods in laboratory medicine. Edited by Rainer Haeckel. Weinheim; New York; Basel; Cambridge : VCH, 1993. ISBN 3-527-30002-3 (Weinheim).

102 National Committee for Clinical Laboratory Standards: Preliminary Evaluation of Quantitative Clinical Laboratory Methods: Tentative Guideline. 2nd ed. Document EP10-T2. Wayne, PA, NCCLS, 1993.

103 National Committee for Clinical Laboratory Standards: Method Comparison and Bias Estimation Using Patient Samples: Approved Guideline. Document EP9-A. Wayne, PA, NCCLS, 1995.

104 Westgard JO. Basic Method Validation manual 1999, 256 pp. ISBN 1-886958-12-2.

105 Clarke WL, Cox D. Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987; 10:622-628.

106 Libeer J-C. Validation of clinical laboratory results: Discussion of essential validation elements. Drug Information Journal 1997; 31: 243-250.

107 Sorto A, Törmä A, Kaihola H-L. Laadunvarmistus kliinisessä laboratoriossa. Sisäisen laadunohjauksen periaatteet. Moodi 1996; Special issue no.5: pp.23. Labquality, Helsinki, Finland. ISBN 951-96219-5-4.

108 Laadunvarmistus kliinisessä kemiassa Yksittäisten sarjojen mittaustekniset kontrollisäännöt. Quality assurance in clinical chemistry. Petersen PH, Dreyer T (original publication in Danish) 1990. Kvalitetssäkring in klinisk kemi. Kallner A (translation in Swedish) 1991. Linko S (translation in Finnish) 1994. NORDKEM Publications Joint project of clinical chemistry, Helsinki, Finland. ISBN 951-47-9988-7, ISSN 0783-2907.

109 Kallner A. Quality specifications based on the uncertainty of measurement. Scand J Clin Lab Invest 1999; 59: 513-516.

110 Kallner A. Uncertainty in measurement, introduction and examples. eJIFCC vol 13 no 1: http://www.ifcc.org/ejifcc/vol13no1/1301200103.htm (May 2002).

111 Örnemark U, Holmgren M. Measurement uncertainty - Surveys about customer´s knowledge, reactions and needs. SP Swedish National Testing and Research Institute. SP Report 2001:07. ISBN 91-7848-849-4.

112 Westgard JO, Bawa N, Ross JW, Lawson NS. Laboratory precision performance. State of the art versus operating specifications that assure the analytical quality required by Clinical Laboratory Improvement Amendments proficiency testing. Arch Pathol Lab Med 1996; 120: 621-625.

113 Dybkaer R. Terminological problems around "accuracy" and "uncertainty". Scand J Clin Lab Invest 1998; 58: 35-46.

114 ISO 5725 (1994-1998), Accuracy (Trueness and precision) of measurement methods and results. Parts 1-6. ISO, Geneva.

115 EURACHEM/CITAC Guide Quantifying Uncertainty in Analytical Measurement, 2nd ed., Eurachem (2000). http://www.eurachem.bam.de (May 2002)

116 Guide to the expression of uncertainty in measurement, "GUM". ISO, Geneva 1993. ISBN 92-67-10188-9.

117 Siloaho M, Elg P, Leppänen E, Loikkanen M, Puukka R, Ruopuro ML, Saarmala H. Ohjeita mittausepävarmuuden arvioimiseksi ja laskemiseksi kliinisen kemian laboratoriossa. Moodi 1997; No.5: 196-203. Labquality, Helsinki, Finland. ISSN 0359-2197.

118 Kouri T, Koskinen P, Leppänen E, Malminiemi OP, Pohja-Nylander P, Pohjavaara S, Puukka R, Siloaho M. Preanalyyttisen mittausepävarmuuden laskeminen. Moodi 2002; No.4: 139-148. Labquality, Helsinki, Finland. ISSN 0359-2197.

119 Kristiansen J. Description of a generally applicable model for the evaluation of uncertainty of measurement in clinical chemistry. Clin Chem Lab Med 2001; 39: 920-931.

120 Johnson J, Patnick J. Achievable standards, Benchmarks for reporting, and Criteria for evaluating cervical pathology. NHSCSP publication no. 1. Cytopathology 2000; 11: 212-241.

121 Smith JE, Winkler RL, Fryback DG. The first Positive: Computing Positive predictive value at the extremes. Ann Intern Med 2000; 132: 804-809.

122 Youden WJ. Index for rating diagnostic test. Cancer 1950; 3: 32-35.

123 Ellison SLR, Gregory S, Hardcastle WA. Quantifying uncertainty in qualitative analysis. Analyst 1998; 123: 1155-1161.

124 Belk WP, Sunderman FW. Survey of the accuracy of clinical analysis in clinical laboratories. Am J Clin Pathol 1947; 17: 853-61.

125 Hamlin WB. The history of evaluation criteria for CAP surveys. Clin Chem 1993;7:1456-1460.

126 Labquality News 1/1998:4. Labquality Days 6.-7.2.1998. Abstracts. Labquality, Helsinki, Finland. ISSN 1237-5616.

127 ISO/IEC (1997) Guide 43-1 Proficiency testing by interlaboratory comparisons, Part 1: Development and operation of proficiency testing schemes, 2nd. edn. International Organization for Standardization, Geneva.

128 Libeer J-C. Role of external quality assurance schemes in assessing and improving quality in medical laboratories. Clin Chim Acta 2001; 309: 173-177.

129 Büttner J, Borth R, Boutwell JH et al.: International Federation of Clinical Chemistry approval recommendation (1983) on quality control in clinical chemistry: V. External quality control. J. Clin Chem Clin Biochem 1983; 21: 885-892.

130 Keenlyside RA, Collins CL, Hancock JS, MariBeth CG, Cohn RD, Menoff AL, Dodd LG, Kurtycz DFI, Hear TL, Baker EL. Do proficiency test results correlate with

the work performance of screeners who screen Papanicolaou smears. Am J Clin Pathol 1999; 112: 769-776.

131 Thomson AH, Watson ID, Wilson JF, Sweeney G, Dawkins CE, Smith BL, Williams J, Capps NE, Toseland PA. An audit of therapeutic drug monitoring service provision by laboratories participating in an external quality assessment scheme. Ther Drug Monit 1998; 3: 248-252.

132 Institute for Reference Materials and Measurements, European Commission, Joint Research Centre, Geel, Belgium. International Measurement Evaluation Programme, IMEP. http://www.irmm.jrc.be and http://imep.ws (November 2002).

133 Örnemark U, Uldall A, Taylor PDP, Van Nevel L, De Bièvre P. Interlaboratory comparisons in support of international comparability and traceability for clinical chemical measurements. EQAnews 1999; 3: 63-64.

134 Örnemark U, Van Nevel L, Taylor PDP, Robouch P, De Bièvre P, Uldall A, Loikkanen M, Müller MM, Libeer J-C, Steensland H, Hellsing K, Squirrell A, Penberthy LA, Sceil D, Tamberg T, Walczyk T, Lam JWH. The IRMM International Measurement Evaluation Programme (IMEP) IMEP-7: Inorganic components in human serum. Accred Qual Assur 1999; 4: 463-472.

135 Lamberth A, Savory J, Moody JR, De Bièvre P, Rosman KJR, Gramlich JW. The IRMM International Evaluation Programme (IMEP) IMEP-4: Trace elements (Li, Cu, Zn) in serum. Accred Qual Assur 1998; 3: 447-458.

136 Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for imprecision and inaccuracy of analytical systems for clinical chemistry. Eur J Clin Chem Clin Biochem 1992; 30: 311-317.

137 Hyltoft Petersen P, Ricos C, Stockl D, Libeer JC, Baadenhuijsen H, Fraser C, Thienpont L. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. Eur J Clin Chem Clin Biochem 1996; 34: 983-999.

138 Hyltoft Petersen P, Fraser CG, Westgard JO, Lytken Larsen M. Analytical goal-setting for monitoring patients when two analytical methods are used. Clin Chem 1992; 38: 2256-60.

139 Fraser CG. Quality specifications in laboratory medicine - current consensus views. Accred Qual Assur 1999; 4: 410-413.

140 Ricós C, Baadenhuijsen H, Libeer JC, Hyltoft Petersen P, Stöckl D, Thienpont L, Fraser CG. External quality assessment: currently used criteria for evaluating performance in European countries. Eur J Clin Chem Clin Biochem 1996; 34: 159-165.

141 Linko S. Quality specifications in laboratory medicine - A literature review. Institute for Reference Materials and Measurements, European Commission - JRC. GE/R/IM/40/01, November 2001.

142 Thienpont L. Quality specifications for reference methods. Scand J Clin Lab Invest 1999; 59: 535-538.

143 Fraser CG, Hyltoft Petersen P. Desirable standards for laboratory tests if they are to fulfill medical needs. Clin Chem 1993; 39: 1447-55.

144 Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiminéz CV, Mincinela J, Perich C, Simón M, Current databases on biological variation: pros, cons and progress, Scand J Clin Lab Invest 1999; 59: 491-500.

145 Fuentes-Arderiu X et al., Pre-metrological (Pre-analytical) Variation of some Biochemical Quantities, Clin Chem Lab Med 1999; 37: 987-989.

146 Cotlove E, Harris EK, Williams GZ. Biological and analytical components of variation in long-term studies of serum constituents in normal subjects. III Physiological and medical implications. Clin Chem 1970; 16: 1028-32.

147 Fraser CG. Desirable performance standards for therapeutic drug monitoring. Clin Chem 1987; 33: 387-9.

148 Moody JP. Biologic variation of serum and salivary lithium. Ther Drug Monit 1999; 21: 97-101.

149 Sohlberg HE. Subject-based reference values. Scand J Clin Lab Invest 1995; Suppl 222: 7-10.

150 Report of a WHO Consultation. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications, Part 1: Diagnosis and Classification of Diabetes Mellitus, Geneva 1999.

151 IFCC/ICSH. Solberg HE. Approved recommendation [1987] on the theory of reference values. Part 1. The concept of reference values. J Clin Chem Clin Biochem 1987; 25: 337-342.

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

153 Solberg HE. Approved recommendations [1987] on the theory of reference values. Part 5. Statistical treatment of collected reference values: determination of reference limits. J Clin Chem Clin Biochem 1987; 25: 645-56.

154 Optimal Analytical Performance for POCT, Fraser CG eJIFCC vol 13 no1: http://www.ifcc.org/ejifcc/vol13no1/1301200106.htm (May 2002).

155 prEN/ISO DIS 15197. Determination of performance criteria for in vitro blood glucose monitoring systems for management of diabetes mellitus. ISO, Geneva, 2000.

156 Kaplan LA. Determination and application of desirable analytical performance goals: the ISO/TC 212 approach. Scand J Clin Lab Invest 1999; 59: 479-482.

157 Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Consensus agreement. Scand J Clin Lab Invest 1999; 59: 585.

158 Fraser CG, Hyltoft Petersen P. Analytical Performance Characteristics should be judged against objective quality specifications. Clin Chem 1999; 45: 321-323.

159 Hartman KE, Nanda K, Hall S, Myers E. Technologic advances for evaluation of cervical cytology: Is newer better? Obstet Gynecol 2001; 56: 765-774.

160 Stockl D, Libeer J-C, Reinauer H, Thienpont LM, De Leenheer AP. Accuracy-based assessment of proficiency testing results with serum from single donations: possibilities and limitations. Clin Chem 1996; 42: 469-470.

161 Food and Drug Administration. Center for Devices and Radiological Health. List of Analytes Waived by FDA.
http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfClia/analyteswaived.cfm
(November 2002)

162 Johnson RN, Baker JR. Error detection and measurement in glucose monitors. Clin Chim Acta 2001; 307:61-67

163 Chan JCN, Wong RYM, Cheung CK, Lam P, Chow CC, Yeung VTF, Kan ECY, Loo KM, Mong MYL, Cockram CS. Accuracy, precision and user-acceptability of self blood blucose monitoring machines. Diabetes Res Clin Practice 1997; 36:91-104